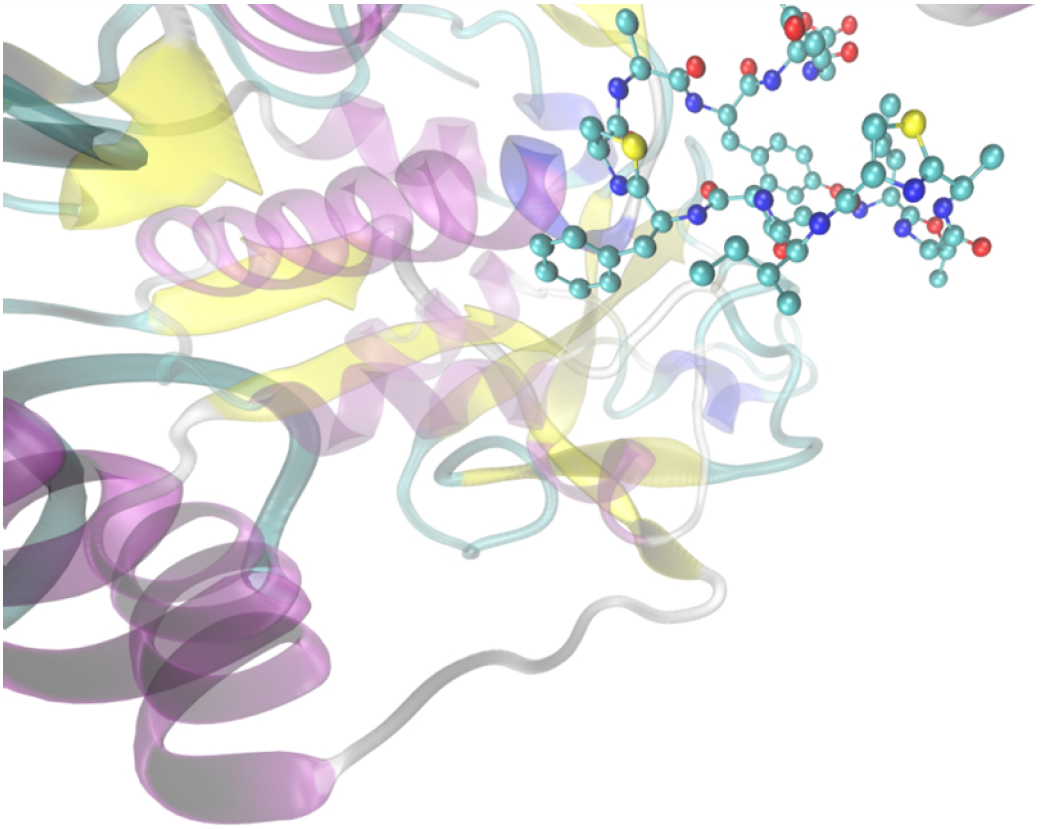


Computational Chemistry Studies on Biological Systems

Anna Bondarenko

Mestrado em Química

Departmento de Química e Bioquímica

2014

Orientadora

Maria João Ramos, Professora Catedrática, Faculdade da Ciências, Universidade do Porto

Coorientadores

Pedro A. Fernandes, Professor Associado, Faculdade da Ciências, Universidade do Porto

Natércia F.Brás, Investigadora, Faculdade da Ciências, Universidade do Porto

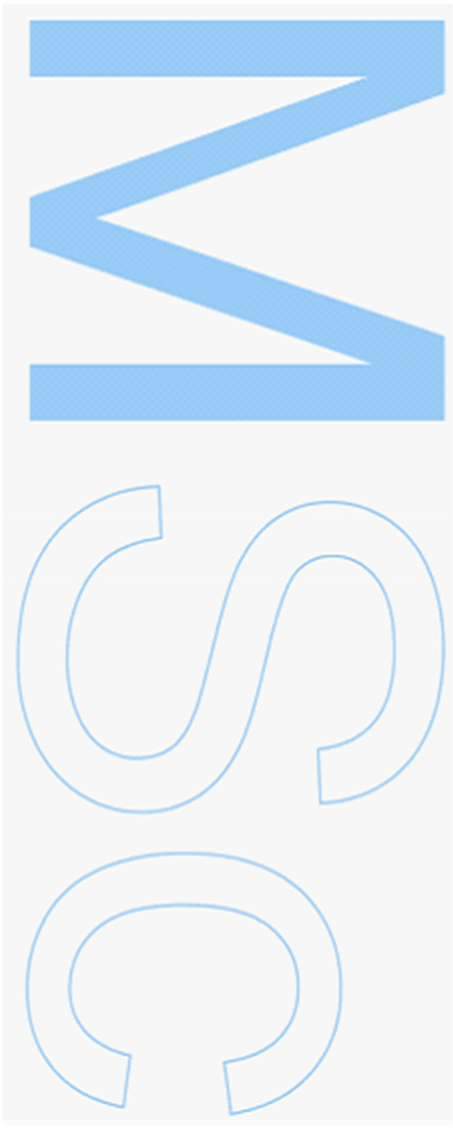
Jasper Knoester, Professor, Faculty of Mathematics and Natural Sciences, University of Groningen

Thomas la Cour Jansen, Assistant Professor, Faculty of Mathematics and Natural Sciences, University of Groningen

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



ACKNOWLEDGEMENTS

This research project would not have been possible without the support of many people. I am gratefully thankful to all the people who was abundantly helpful and offered invaluable assistance, support and guidance.

First and foremost, I would like to give special thanks to the Erasmus Mundus programme of the European Union, with the support of which is done the project. I am thankful for the opportunity to receive the international European master degree in Theoretical Chemistry and Computational Modelling, to have the experience of doing the project in two countries.

I wish to express my acknowledges for ability to carry this project at the research group of Theoretical Chemistry, Department of Chemistry, Faculty of Science, University of Porto, Porto. In particular, my deepest gratitude is to my supervisors Maria João Ramos, Pedro Alexandrino Fernandes and Natércia Bras.

I gratefully acknowledge to another research group, for opportunity to carry the part of the research project – the group of Theory of Condensed Matter of Zernike Institute for Advanced Materials, Faculty of Mathematics and Natural Sciences, University of Groningen, Groningen. I am highly thankful to the coordinators of the project – Thomas la Cour Jansen and Jasper Knoester.

I would like to give thanks to the experimental groups for providing the experimental data. Marcel Jaspars are deeply acknowledged for providing X-ray structures of enzymes. Andrei Tokmakoff and Carlos Baiz are gratefully acknowledged for providing the experimental FTIR and two-dimensional infrared spectra data for proteins.

Finally, I wish to express my heartfelt thanks to my beloved family, supporting and encouraging me through the duration of my studies, my friends for their help and wishes for the successful completion of this project.

ABSTRACT

The current research project was developed in two lines within the main framework of the application of computational chemistry approaches on studies of biological systems. The first line, carried in the group of Theoretical Chemistry of the University of Porto, is related to the area of drug design and includes two parts of the project. The first part aimed at providing insights for the design of novel drugs for treatment of hypertension. Here, the model of the target angiotensin II receptor type 1 (AT1) receptor, which is involved in the renin-angiotensin-aldosterone-system (RAAS) cascade, was modelled by homology modelling approach. The binding poses of the natural substrate Angiotensin II, peptide hormone that causes vasoconstriction, to the created model of the target were identified by molecular docking study. The molecular docking protocol was also employed to dock several AT1 antagonists to the modelled target. The second part aimed to perform preparation of the enzyme macrocyclase PatG (PatGmac) system for the study of its catalytic mechanism. The system was created on the base of X-ray structure and fully optimised by the hybrid QM/MM ONIOM method. In future, it would be useful for production of potential anticancer molecules "patellamides", linear peptides cyclised by this enzyme.

The second line of the project, carried in the group of Theory of Condensed Matter of the University of Groningen, is related to the study of protein conformational changes by means of two-dimensional infrared spectroscopy (2DIR). Theoretical methods capable of detailed simulation of two-dimensional spectra of full proteins from molecular dynamics trajectories have been applied for the proteins to test the method on big biological systems.

Key words: Drug Design, hypertension, RAAS, AT1 receptor, angiotensin II, AT1 antagonists, homology modelling, molecular docking, PatGmac, patellamides, ONIOM QM/MM, 2DIR spectroscopy, Molecular Dynamics

RESUMO

O presente projeto de investigação foi desenvolvido em duas linhas, tendo por base abordagens de química teórica e computacional em estudos de sistemas biológicos. A primeira componente, realizada no grupo de Química Teórica da Universidade do Porto, está relacionada com um estudo de desenho de fármacos e divide-se em duas partes deste projeto. Na primeira parte pretendeu-se a aquisição de conhecimentos para a concepção de novos fármacos para o tratamento da hipertensão. Inicialmente modelou-se o receptor transmembranar da angiotensina II, subtipo 1 (AT1), o qual está envolvido na cascata sistema-renina-angiotensina-aldosterona (RAAS), utilizando o método de modelação por homologia. Utilizou-se o método de atracagem molecular para se identificaram as diferentes conformações de ligação do substrato natural angiotensina II, hormona peptídica que provoca vasoconstrição, ao modelo do receptor AT1 criado por modelação por homologia. O mesmo protocolo de atracagem molecular foi posteriormente utilizado para atracar vários antagonistas do receptor AT1 (comercialmente utilizados) ao alvo modelado.

Na segunda parte realizou-se a preparação do sistema da enzima macrociclase PatG (PatGmac) para o futuro estudo do seu mecanismo catalítico. O sistema foi criado a partir da estrutura de cristalografia de raio-X, e de seguida foi optimizado pelo método híbrido QM/MM ONIOM. Este procedimento é necessário para o estudo do mecanismo catalítico, o qual será muito útil para a produção de moléculas "patelamidas", péptidos cíclicos produzidos por esta enzima, que actuem como potenciais fármacos anti-cancerígenos.

A segunda linha do projeto, realizado no grupo de Teoria da Matéria Condensada da Universidade de Groningen, está relacionada com o estudo de mudanças conformacionais de proteínas por meio de espectroscopia de infravermelho bidimensional (2DIR). Foram aplicados métodos teóricos capazes de simular detalhadamente os espectros bidimensionais de proteínas completas através trajectórias de dinâmica molecular. O método foi testado em proteínas pertencentes a grandes sistemas biológicos.

Palavras-chave: Desenho de fármacos, Hipertensão, RAAS, Receptor AT1, Angiotensina II, Antagonistas do AT1, Modelação por Homologia, Atracagem Molecular, PatGmac, Patelamidas, ONIOM QM/MM, 2DIR Espectroscopia, Dinâmica Molecular

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

2DIR	Two-dimensional infrared spectroscopy
7TM	seven-transmembrane domain
ACE	angiotensin-converting enzyme
Ang-I	angiotensin I
Ang-II	angiotensin II
ARB	Angiotensin II receptor blocker
AT1	angiotensin II receptor type 1
bRh	bovine Rhodopsin
Con A	Concanavalin A
DOPE	Discrete Optimized Protein Energy
E-value	Expected value
FTIR	Fourier Transformed Infrared spectroscopy
GA	Genetic Algorithm
GLDP	glycin dipeptide
GPCR	G protein-coupled receptor
HM	homology modelling
LGA	Lamarckian Genetic Algorithm
Lys	Lysozyme
Mb	Myoglobin
NMA	*N*-methyl acetamide
NMR	Nuclear Magnetic Resonance
pat A	patellamide A
pat B	patellamide B
pat C	patellamide C
pat D	patellamide D
pat E	patellamide E
pat G	patellamide G
PatGmac	macrocyclase PatG
PDB	Protein Database
RAAS	renin-angiotensin-aldosterone system
RMSD	root-mean square deviation
RNse A	Ribonuclease A
SA	Monte Carlo Simulated Annealing
Sar	sarcosine
TCC	transition charge coupling
TDC	transition dipole coupling
VS	Virtual Screening

I. INSIGHTS FOR THE DESIGN OF NOVEL DRUGS FOR HYPERTENSION

1. INTRODUCTION

Hypertension, or high blood pressure, is a chronic condition recognized as one of the leading causes for human morbidity and mortality worldwide through its effects on target organs like heart, brain and kidneys. It is a major risk factor for myocardial infarction, congestive heart failure, stroke, and end-stage renal disease all of which convey risk of significant morbidity and mortality [1].

Despite the huge problem of the disease and the existence of several different classes of antihypertensive pharmacological drugs, the incidence of hypertension and related cardiovascular diseases remains steadily on the rise, and relatively few patients achieve the targeted blood pressure level. Moreover, majority of available pharmacological agents for hypertension present severe side effects to the organism.

The renin-angiotensin-aldosterone system (RAAS) plays an important role not only in the control of blood pressure, but in the pathogenesis of diabetes and kidney diseases. While it has been difficult to demonstrate *in vivo* activation of the RAAS in early or established hypertension in humans, there is no doubt that inhibition of the RAAS is effective in lowering blood pressure in patients with primary hypertension. The results of multiple clinical trials demonstrate that blocking the RAAS through the targets of angiotensin-converting enzyme (ACE) or angiotensin II receptor type 1 (AT1), which are parts of the proteolytic cascade of the RAAS, with the corresponding inhibitors not only lower blood pressure, but also reduces cardiovascular events and total morbidity and mortality [1, 2]. Thus, it makes the study of this system highly important, and design of new blocking agents to this cascade is an attractive therapeutic strategy.

The aim of the project was to provide insights for the design of novel drugs for hypertension involving theoretical and computational calculations. The study has been performed focusing on creating a model of the AT1 receptor, identifying the binding poses of Angiotensin II and several AT1 antagonists. Homology modelling, force-field minimization, and molecular docking calculations were used during the study.

2. RENIN-ANGIOTENSIN-ALDOSTERONE-SYSTEM (RAAS) IN HYPERTENSION

2.1 GPCR receptors

G protein-coupled receptors (or GPCRs), also known as seven-transmembrane domain receptors (7TM receptors), are integral membrane proteins that contain seven membrane-spanning helices. As the name suggests they are coupled to heterotrimeric G proteins on the intracellular side of the membrane. Upon ligand binding, the GPCR undergoes a conformational change that is transmitted to the G protein causing activation. Further signal transduction depends on the type of G protein.

GPCRs are the largest and most diverse group of membrane receptors in eukaryotes. They are activated by a wide variety of ligands, olfactory stimulants, peptides, hormones, neurotransmitters and light. They are grouped into 6 classes based on sequence homology and functional similarity:

Class A (or 1) (Rhodopsin-like);

Class B (or 2) (Secretin receptor family);

Class C (or 3) (Metabotropic glutamate/pheromone);

Class D (or 4) (Fungal mating pheromone receptors);

Class E (or 5) (Cyclic AMP receptors);

Class F (or 6) (Frizzled/Smoothened).

GPCRs are the target of around half of all modern medicinal drugs. Their expression on the cell surface makes them readily accessible to hydrophilic drugs and their non-uniform expression provides selectivity in activating or blocking physiological events. Agonists and antagonists of GPCRs are used in the treatment of disease in every organ system.

GPCRs are the targets of almost 50% of all drugs currently available on the market [3]. Class A GPCRs (or Rhodopsin-like) share the same basic topology of seven transmembrane helices connected by three extracellular and three intracellular loops. The receptor is capped with an extracellular amino terminus and intracellular carboxyl terminus (Fig. 1).

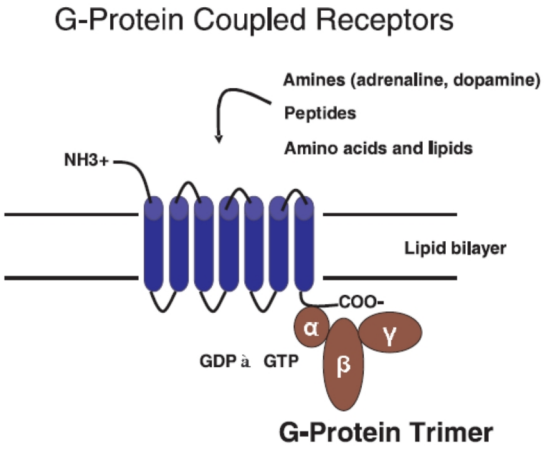


Fig. 1 – Schematic representation of the class A GPCR topology. This picture shows an "activated" GPCR embedded in a lipid bilayer. The C-terminus of the receptor is associating with the G-protein trimer.

2.2. AT1 receptor

The angiotensin II type I (AT1) receptor is a class A GPCR (Fig. 1) and is involved in the RAAS (Fig. 2).

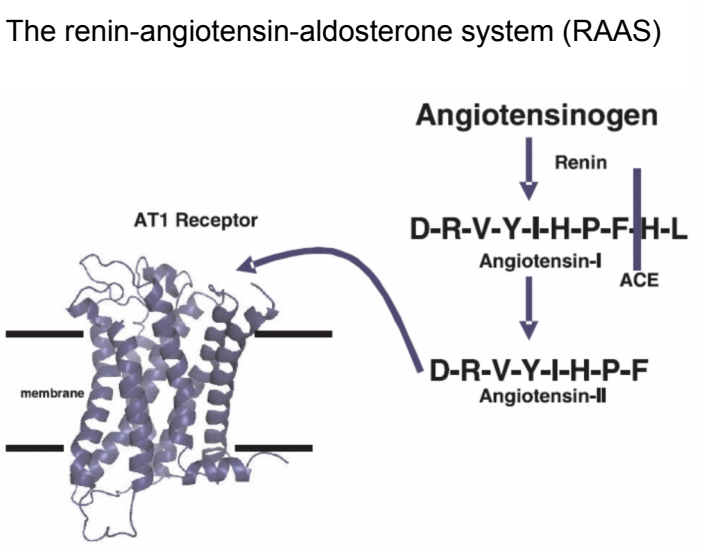


Fig. 2 – A diagram illustrating the RAAS cascade, beginning with angiotensinogen and ending with the binding of angiotensin-II to the AT1 receptor.

The RAAS plays a key role in blood pressure regulation and electrolyte homeostasis, and it is involved in the pathogenesis of hypertension and renal diseases. The RAAS system is a proteolytic cascade that starts with the release of the enzyme renin by the juxtaglomerular cells located in the kidney. Renin binds a plasma protein named angiotensinogen (452 amino acids) and hydrolyses the peptidic bond between the 10^{th} and 11^{th} positions, releasing a decapeptide known as angiotensin I (Ang-I). Ang-I remains biologically inactive until it is further cleaved into the octapeptide angiotensin II (Ang-II) by the metalloprotease angiotensin-converting enzyme (ACE) [4]. Ang-II is the main effector

hormone of the RAAS. Ang-II subsequently binds and activates a transmembrane receptor AT1 located in the cardiovascular, neuronal, endocrine and hepatic cells [5]. AT1 activation – the last binding of Ang-II to AT1 receptor – induces cardiovascular events such as vasoconstriction, inflammatory response, aldosterone release and salt retention [6, 7]. Many antagonists have been designed to block this interaction in order to treat clinical cases of hypertension [8].

The AT1 receptor, a 359-aminoacid protein, is widespread in organs and tissues, but is found predominately in vascular and myocardial tissue, the liver, the adrenal cortex and some areas of the brain [9, 10]. The design and development of AT1 blockers could be useful as specific remedies for many pathologies like hypertension, heart and kidney failures. Therefore, a knowledge of the 3D structure of the AT1 receptor could be of great help in this task.

The 3D-structure of a protein is usually obtained experimentally and the main techniques for this purpose are X-ray crystallography and Nuclear Magnetic Resonance (NMR). Unfortunately, the crystallization of almost all GPCRs is still an unresolved problem [11].

Given the high-level of interest these proteins attract, it is not surprising that predictive methods to derive information about their 3D-structure are essential. An alternative approach to experimental methods is to build a molecular model of the protein through the homology modelling (HM), [12] in which the target protein is constructed starting from the experimentally known 3D-structure of related proteins.

The first high-resolution structure of a protein related to GPCRs was published for bovine rhodopsin [13], thus more reliable HM approaches for all the GPCRs begun to be affordable. Recently, some other related proteins were resolved [14, 15, 16]. Below in the section 3.2.1, all the available crystallographic structures would be reported.

2.3. Angiotensin II receptor blockers (ARBs)

The initial target for developing compounds that could inhibit the renin-angiotensin pathway was the angiotensin II octapeptide (Asp1-Arg2-Val3-Tyr4-Ile5-His6-Pro7-Phe8). Ang-II plays a key role in the regulation of cardiovascular homeostasis. Acting on both the “content” and the “container”, Ang-II regulates blood volume and vascular resistance. The wide spectrum of Ang-II target tissues includes the adrenals, kidney, brain, pituitary gland, vascular smooth muscle, and the sympathetic nervous system.

Angiotensin II receptor blocker (ARBs) are designed to displace angiotensin II from the AT1 receptor and produce their blood pressure–lowering effects by antagonizing angiotensin II–induced vasoconstriction, aldosterone release, catecholamine release, arginine vasopressin release, water intake, and hypertrophic response [17].

Efforts to develop angiotensin II receptor antagonists began in the early 1970s and focused on peptide-based analogues of the natural agonist. The prototypical compound that resulted from these studies was saralasin, an octapeptide in which the Asp1 and Phe8 residues of angiotensin II were replaced with Sar (sarcosine, N-methylglycine) and Ile, respectively. Saralasin, as well as other peptide analogues demonstrated the ability to reduce blood pressure. However, these compounds lacked oral bioavailability and has to be administered intravenously [18], and it has behaved as a partial agonist [19].

More recent efforts have used peptide mimetics to pass over these inherent problems with peptide-based antagonists. In early 1980s it was noted that a series of imidazole-5-acetic acid derivatives decreased blood pressure responses to Ang-II in rats. Two compounds, S-8307 and S-8308 [20], were later found to be highly specific and promising non-peptide Ang-II receptor antagonists, but afterwards it was seen that their structures would have to mimic more closely the pharmacopoeia of Ang-II. Structural modifications were made and the orally active, potent and selective non-peptide AT1 receptor blocker losartan was developed. As a result of these efforts in 1995, losartan first to be approved for clinical use in the United States and since then seven additional ARBs have been approved [21, 22] (Table 1).

Losartan, developed to mimic a natural peptide Ang-II, shares common structural features with it. The imidazole ring of losartan correlated with the imidazole side chain of the His6 residue, and the n-butyl group of losartan correlated with the hydrocarbon side chain of the Ile5 residue. Valsartan, irbesartan, telmisartan, candesartan, olmesartan, and azilsartan are biphenyl analogs of losartan.

Valsartan represents a nonheterocyclic AT1 receptor selective antagonist in which the imidazole of losartan has been replaced by an acylated amino acid [23].

Irbesartan is a potent AT1 receptor antagonist that incorporates an imidazolinone ring in which a carbonyl group functions as a hydrogen bond acceptor in place of the hydroxymethyl group of losartan [24].

The benzimidazole, candesartan cilexetil, developed at Takeda Chemical Industries, is an ester carbonate prodrug that is rapidly converted to the corresponding 7-carboxylic acid, **candesartan** *in vivo* [25]. Candesartan is a potent, long acting selective AT1 receptor, competitive insurmountable antagonist [26]. The carboxyl group at the benzimidazole ring plays an important role in the interaction of candesartan with AT1 receptor [27].

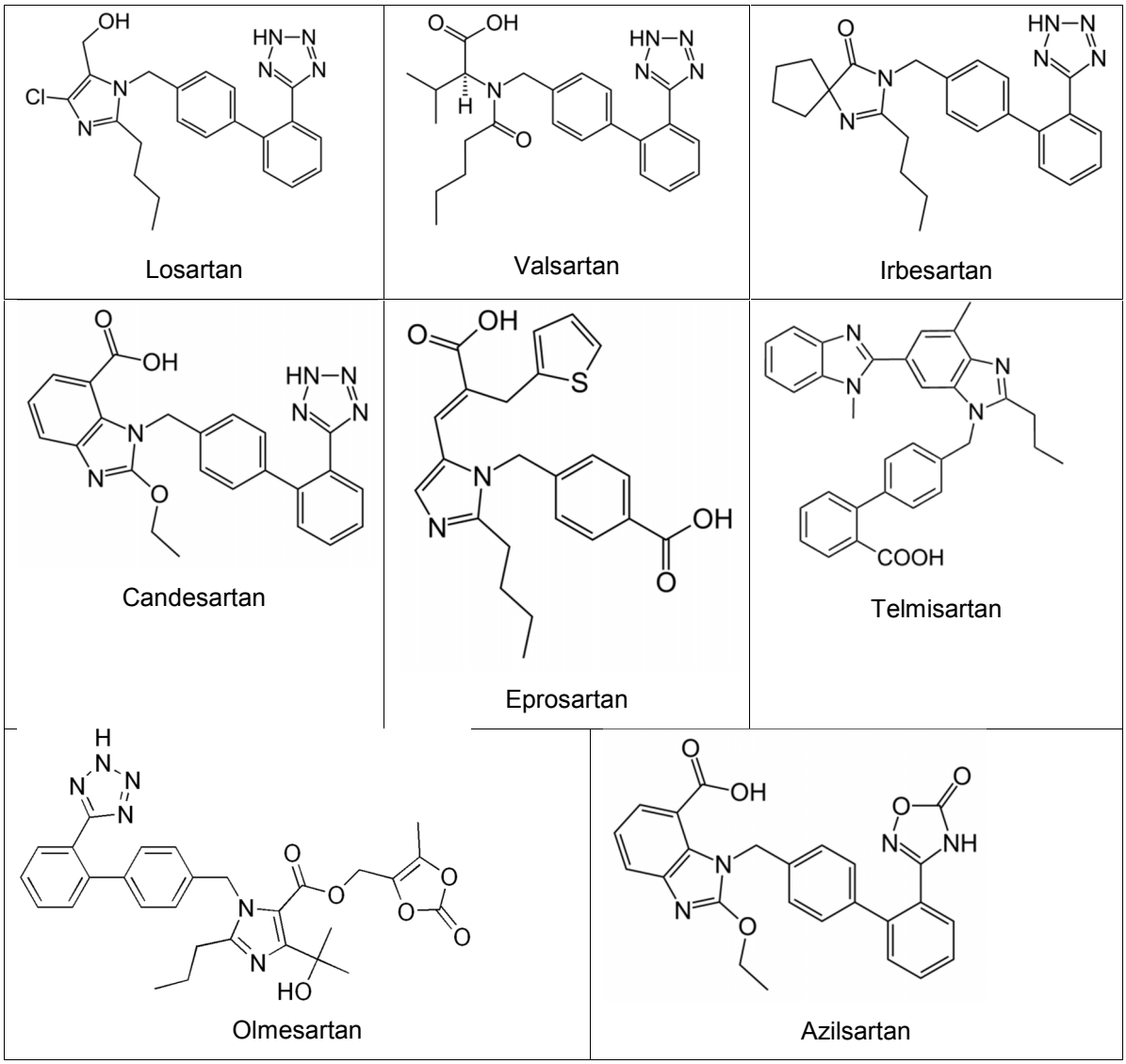


Table 1 – Structures of AT1 receptor nonpeptidic antagonists.

3. MODELLING THE AT1 RECEPTOR

3.1. Homology modelling methodology

Generally, the study of a protein's function requires previous knowledge of its three-dimensional structure [28, 29], which is ultimately determined by protein sequence [30]. Protein structure determination using experimental methods, such as X-ray crystallography or NMR spectroscopy is time consuming and not successful with all proteins, especially membrane proteins [31]. The huge gap between the number of available sequences and experimentally solved protein structures could possibly be resolved by computational methods.

The homology modelling method predicts the three-dimensional structure of a given protein sequence based primarily on its sequence similarity to one or more proteins of known structures. It is also called comparative modelling or template-based modelling.

Homology modelling methods use the fact that evolutionary related proteins share a similar structure [32, 33]. A protein structure is always of great assistance in the study of interactions with ligands within pharmaceutical industry in structure-based drug discovery and drug design. Homology modelling can provide the molecular biologists and biochemists with "low-resolution" structures, which will contain sufficient information about the spatial arrangement of important residues in the protein and that may guide the design of new experiments.

The protein under the query with unknown 3D structure is called the *target*. Whereas, the protein or proteins with similar sequence and known 3D structure, that would be used to predict the structure of the target protein is known as *template* protein or proteins.

In practice, homology modelling is a multistep process that can be summarized in four steps: 1) template selection; 2) target-template alignment; 3) model construction and 4) model assessment.

Step 1. Template selection

The sequence of similarity can be searched using BLAST or Psi blast or fold recognition methods and align with the known structures in Protein Database (PDB). PDB is the largest database that contains only experimentally resolved structures. BLAST allows comparing a query sequence with a database, such as PDB and identifying the best sequence that shares a high degree of similarity. The sequence of similarity of each line is summarised with its E-value (Expected value). E-values closer to zero, have high degree of similarity. The E-value describes the number of hits one can “expect” when searching through a database of a particular size. The sequences that fall under safe zone are expected to be getting good structure than twilight zone and midnight zone.

Step 2. Target-template alignment

After identifying one or more possible template, alignment is performed. Sometimes it is difficult to align two sequences that have low percentage of identity. Such cases, one can use other sequences from homologous proteins to solve this problem. Multiple Sequence Alignment programs such as CLUSTALW [55] align sequences by insertions and deletions. Alignment correction is the critical step in Homology Modelling, otherwise which in turn creates a defective model.

Step 3. Model construction (Backbone generation)

a. Backbone generation

The backbone generation from the aligned regions can be done using modelling tools such as Modeller [34]. The actual experimentally determined structures contain

manual errors due to poor electron density in the map. Therefore a good model has to be chosen with less number of errors.

Modeller is a program for comparative protein structure modelling by satisfaction of spatial restraints [35]. It can be described as “Modelling by satisfaction of restraints” uses a set of restraints derived from an alignment and the model is obtained by minimization of these restraints. These restraints can be from related protein structures or NMR experiments. User gives an alignment of sequences to be modelled with known structures. Modeller calculates a model with all heavy atoms. It also performs comparison of protein structures or sequences, clustering of proteins, searching of sequence databases.

The highest sequence identity, observed for CXCR4 Chemokine receptor (34%), is still small for basic safe homology modelling. Therefore, two more sophisticated approaches were used for modelling: iterative modelling and advanced modelling.

Iterative modelling was introduced to overcome the problem in homology modelling caused by limitation of errors in the alignment of a modelled sequence with related proteins of known three-dimensional structure. Thus, it have been developed an automated method that optimizes both the alignment and the model implied by it [36]. This task is achieved by a genetic algorithm protocol that starts with a set of initial alignments and then iterates through re-alignment, model building and model assessment to optimize a model assessment score.

Advanced modelling, based on multiple templates, utilizes information from various templates, and therefore “naturally” increases the accuracy of Homology Modelling, presumably since it better captures the variability and divergence of natural structures.

b. Loop modelling

In most cases, alignment between model and template sequence contain gaps. By means of insertions and deletions with some conformational changes to the backbone it can be modelled, although it rarely happens to secondary structures. Therefore, it is safe to shift the insertion and deletions of the alignment, out of helices or strands and placing them in loops or coils. But this loop conformational change is difficult to predict due to many reasons like:

1. Surface loops tend to be involved in crystal contacts, leading to a significant conformational change between template and target.

2. The interchange of the side chains can lead to change in the orientation and spatial arrangement, especially when it is an interchange between small and a bulky group.

3. Proline and glycine are an exception when a Ramachandran plot is considered. Proline has a restriction in the plot due to its 5 membered ring, whereas glycine has a

hydrogen atom as its side chain which is very difficult to predict from the plot. This makes it difficult for detect mutations that have happened to loop residue from/to either glycine or proline.

In Modeller, the loop modelling is based on the loop refinement – optimization method for loop modelling. This method relies on a scoring function and optimization schedule adapted for loop Modelling. It is used automatically to refine comparative models if one uses the command from *loopmodel* class rather than *automodel*.

c. Side chain modelling

Proteins that are structurally similar, have similar χ_1–angles (i.e., the torsion about Cα-Cβ bond. In such cases, copying conserved residues entirely from the template to the model will result in higher accuracy than copying the backbone or re-predicting side chains. Side chain conformations are partially knowledge based which uses libraries of rotamers extracted from high resolution X-ray structures. To build a position-specific rotamer library, one can take high-resolution protein structures and collects all stretches of three to seven residues (method dependant) with a given amino acid at the center. Prediction accuracy is usually quite high for residues in the hydrophobic core, it is much lower for surface residues.

There are two reasons for this:

1. Flexible side chains on the surface tend to adopt multiple conformations, which are additionally influenced by crystal contacts.

2. Energy functions used to score rotamers can easily handle hydrophobic packing in the core (Van der Waals interactions), but are not accurate enough to get complicated electrostatic interactions on the surface.

d. Model optimization

Sometimes the rotamers are predicted based on incorrect backbone or incorrect prediction. Such cases Modelling programs either restrain the atom positions and/or apply only a few hundred steps of energy minimization to get an accurate value. This accuracy can be achieved by 2 ways:

1. Quantum force field: To handle large molecules efficiently force field can be used, energies are therefore normally expressed as a function of the positions of the atomic nuclei only. Van der Waals forces are, for example, so difficult to treat, that they must often be completely omitted. While providing more accurate electrostatics, the overall precision achieved is still about the same as in the classical force fields.

2. Self-parametrizing force fields: The precision of a force field depends to a large extent on its parameters (e.g., van der Waals radii, atomic charges). These parameters are usually obtained from quantum chemical calculations on small molecules and fitting to experimental data, following elaborate rules [37]. By applying the force field to proteins,

one implicitly assumes that a peptide chain is just the sum of its individual small molecule building blocks—the amino acids.

Step 4. Model assessment

The obtained models may contain errors, which mainly depend upon two values:

1. The percentage identity between the template and the target. If the value is >90%, then accuracy can be compared to crystallography, except for a few individual side chains. In this case sequence similarities are considered accurate enough for drug discovery applications. If its value ranges between 50-90%, the root-mean square deviation (RMSD) error can be as large as 1.5 Å, with considerably more errors. Those between 25 and 50% identities can be helpful in designing of mutagenesis experiments. If the value is <25% the alignment turns out to be difficult for Homology Modelling, often leading to quite larger errors [38, 39].

2. The number of errors in the template. Errors in a model become less of a problem if they can be localized. Therefore, an essential step in the Homology Modelling process is the verification of the model. The errors can be estimated by calculating the model's energy based on a force field. This method checks to see if the bond lengths and angles are in a normal range. However, this method cannot judge if the model is correctly folded. The 3D distribution functions can also easily identify misfolded proteins and are good indicators of local model building problems.

The prediction and assessment of protein structures can be enabled by accurate free energy function [40]. The native structure generally has the lowest free energy of all states under the native conditions. In principle, the free energy surface of a protein can be derived by thoroughly sampling the potential energy surface defined by a molecular mechanics force field [41]. However, this approach is computationally prohibitive and may be further limited by errors in potential energy functions. Instead of relying on free energy, an alternative approach is to construct a scoring function whose global minimum also corresponds to the native structure from a sample of native structures of different sequences [42] deposited in the Protein Data Bank. Due to its dependence on known protein structures, such a scoring function is often termed a knowledge-based or statistical potential. Such statistical potential was derived and implemented to the MODELLER software, and is named as Discrete Optimized Protein Energy, or DOPE [43]. The DOPE model score was designed for selecting the best structure from a collection of models built by MODELLER software. For example, when the multiple models are built, one can select the model that returns the lowest DOPE score.

Another way to assess the models is a GA341 assessment score, a score for the reliability of a model, derived from statistical potentials [44]. GA341 scores always range from 0.0 (worst) to 1.0 (native-like). A model is predicted to be reliable when the model

score is higher than a pre-specified cutoff (0.7). A reliable model has a probability of the correct fold that is larger than 95%. However GA341 is not as good as DOPE at distinguishing 'good' models from 'bad' models.

3.2. Homology modelling of the AT1 receptor

Below, the structures that are used as templates for the target, found in the PDB, are discussed.

Several attempts have been made to model the human AT1 receptor structure using bacteriorhodopsin 3D structure [45, 46, 47] as a template. Because bacteriorhodopsin is not class A GPCR, direct use of its structure to model the TM-domain structure of the AT1 receptor is limited. Recent availability of bovine Rhodopsin (bRh) X-ray structures made it possible to model the AT1 receptor. Below are reviewed these X-ray structures.

Family of Bovine Rhodopsin receptor

In the year 2000, Palczewsky et al. reported the first high resolution structure of bovine rhodopsin [13]. The structure with a 2.8 Å resolution can be found in the Protein Data Bank (PDB) [48] with PDB ID 1F88 (Fig. 3a). This surely was a milestone towards understanding of the structure of the other GPCRs and for the evolution of the HM procedures.

Other high-resolution models of bRh, obtained using X-ray diffraction, have been successively deposited in the PDB by Teller et al. – PDB ID 1HZX [49] (Fig. 3b), and by Li et al. – PDB ID 1GZM [50] (Fig. 3c). The structures have comparable resolution to PDB ID 1F88.

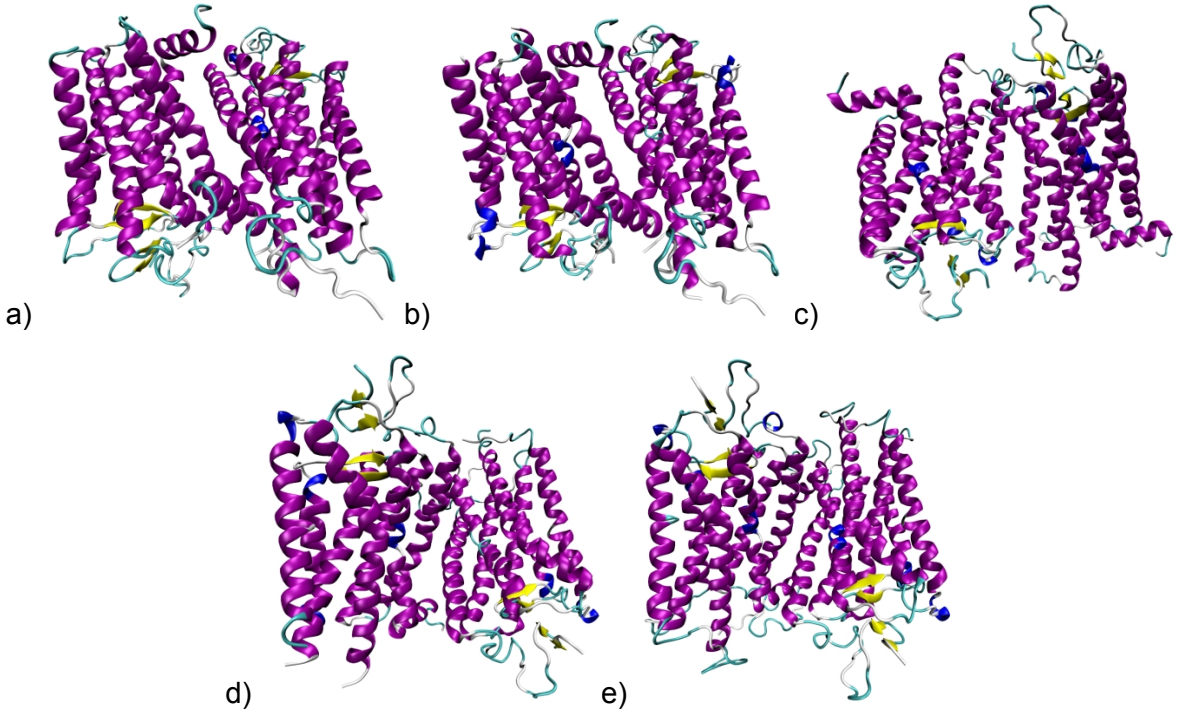


Fig. 3 – Bovine rhodopsin receptor: a) 1F88, b) 1HZX, c) 1GZM, d) 1L9H, e) 1U19

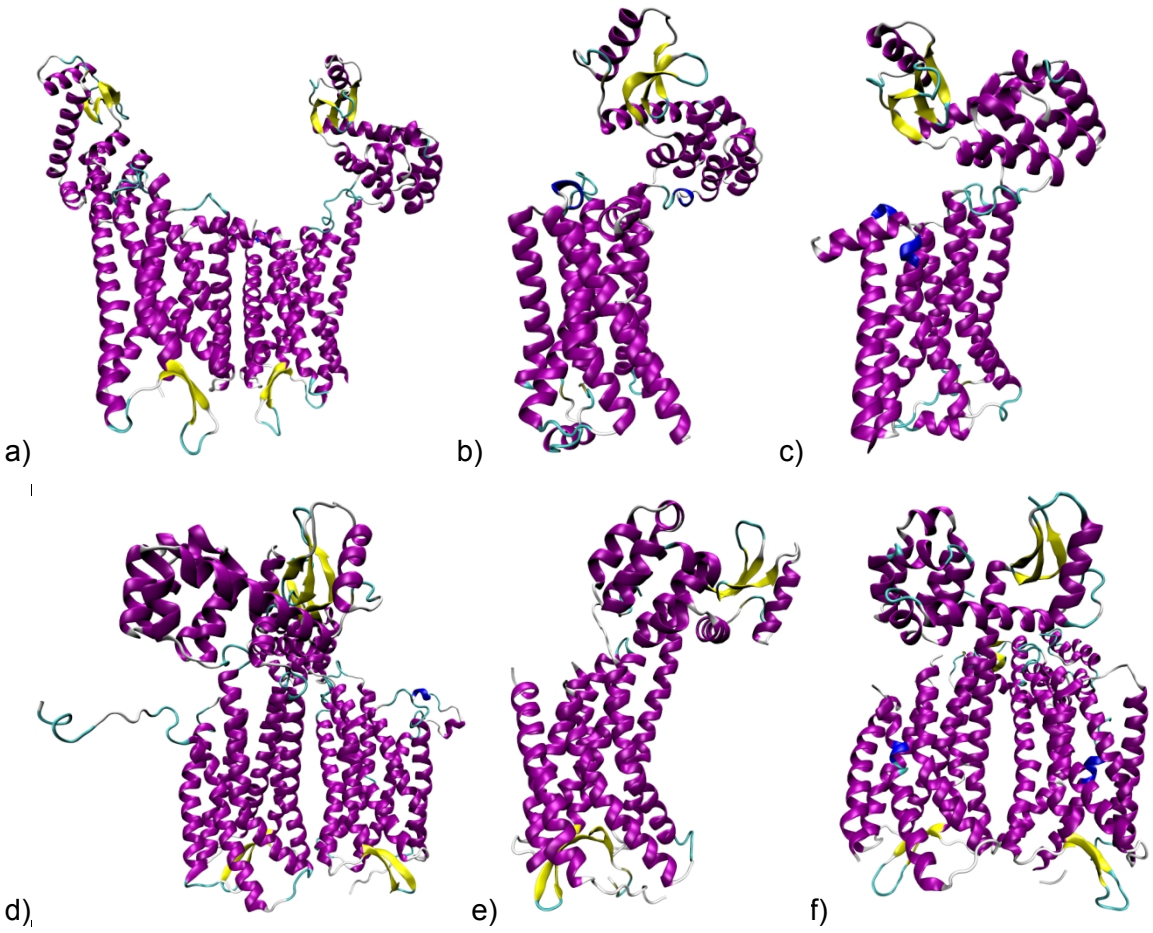


Fig. 4 – Human kappa opioid receptor: a) 4DJH; human β_2-adrenergic receptor: b) 2RH1, c) 3NY9; CXCR4 receptor: d) 3ODU, e) 3OE0, f) 3OE9.

Similarly, another structure was crystallised and published by Okada et al. – PDB ID 1L9H [51] (Fig. 3d). This structure possesses the resolution of 2.6 Å, moreover, it shows that seven water molecules bind some of the highly conserved residues among rhodopsin-like GPCRs and are able to mediate intramolecular interactions between the seven TM domains. Therefore this 3D structure is to be considered of particular importance for the GPCR HM development.

More recently, a new structure (1U19 [52], Fig. 3e) has been available in the PDB. It was crystallised under different conditions, allowing an improvement of the quality of the HM template structure because the resolution was reduced to 2.2 Å.

Family of human kappa opioid receptor

In 2012, another structure from the family of GPCR receptors – human kappa opioid receptor (4DJH [53], Fig. 4a) has been resolved and deposited to the PDB.

Family of human β_2-adrenergic receptor

The crystal structure of one of the most extensively studied receptors – human β_2-adrenergic receptor (2RH1 [14], Fig. 3f) – was reported by Cherezov et al. The structure provides a high-resolution (2.4 Å) view of a human G protein-coupled receptor bound to a diffusible ligand. More recently, another 3D structure of the same receptor in complex with two inverse agonists (3NY9 [15]) was resolved by Wacker et al.

Family of CXCR4 receptor

The branch of γ-GPCR includes available crystallographic structures. In particular, the CXCR4 receptor retains most of the common GPCR structural characteristics, whereas the fourth transmembrane domain (TM4) differs in length as it forms an extra α-helix turn near extracellular part. Wu et al. [16] reported about independent crystal structures of CXCR4 bound to an antagonist small molecule IT1t and a cyclic peptide CVX15 at 2.5 to 3.1 angstrom resolution (3ODU, 3OE0, 3OE9).

The sequences that are necessary for sequence alignment were obtained from the SwissProt Database [54].

All possible templates, used in this study, are summarized below in the Table 2. The table includes the name of the receptor, PDB code, sequence identity code, and X-Ray resolution.

The initial multiple sequence alignment was performed using ClustalW as provided on the EBI webserver [55]. The program compares the query sequence to the template sequences of known structures.

Template Receptor	Sequence (UniProt)	PDB code	X-Ray Resolution
1. Bovine Rhodopsin	P02699	1F88	2.80 Å
		1U19	2.20 Å
		1GZM	2.65 Å
		1HZX	2.80 Å
		1L9H	2.60 Å
2. Kappa-type opioid receptor	P41145	4DJH	2.90 Å
3. Beta-2 adrenergic receptor	P07550	2RH1	2.40 Å
		3NY9	2.84 Å
4. CXCR4 Chemokine receptor	P61073	3ODU	2.50 Å
		3OE0	2.90 Å
		3OE9	3.10 Å

Table 2. Available target templates

In the Table 3 below, the results of multiple alignments are presented in the order from the highest to lowest sequence identity percentage. One is interested in the highest values – the highest similarity between the template and the target proteins. The highest sequence identity is observed for CXCR4 Chemokine receptor – 34%. This is followed by Kappa-type opioid receptor (29%), Beta-2 adrenergic receptor (21%), and Bovine Rhodopsin with 20%.

Template Receptor	Sequence identity	PDB code	X-Ray Resolution
1. CXCR4 Chemokine receptor	34%	3ODU	2.50 Å
		3OE0	2.90 Å
		3OE9	3.10 Å
2. K-opioid receptor	29%	4DJH	2.90 Å
3. β_2 adrenergic receptor	21%	2RH1	2.40 Å
		3NY9	2.84 Å
4. Bovine Rhodopsin receptor	20%	1F88	2.80 Å
		1U19	2.20 Å
		1GZM	2.65 Å
		1HZX	2.80 Å
		1L9H	2.60 Å

Table 3. The percentage of the sequence identity for each template protein.

In the Fig. 5 the multiple sequence alignment between the target receptor AT1 and the template receptors are presented. The initial alignment was manually adjusted to best satisfy the position of residues conserved throughout class A GPCR's [56]. Any gaps were avoided in the trans-membrane helices and instead placed in the loop and termini regions.

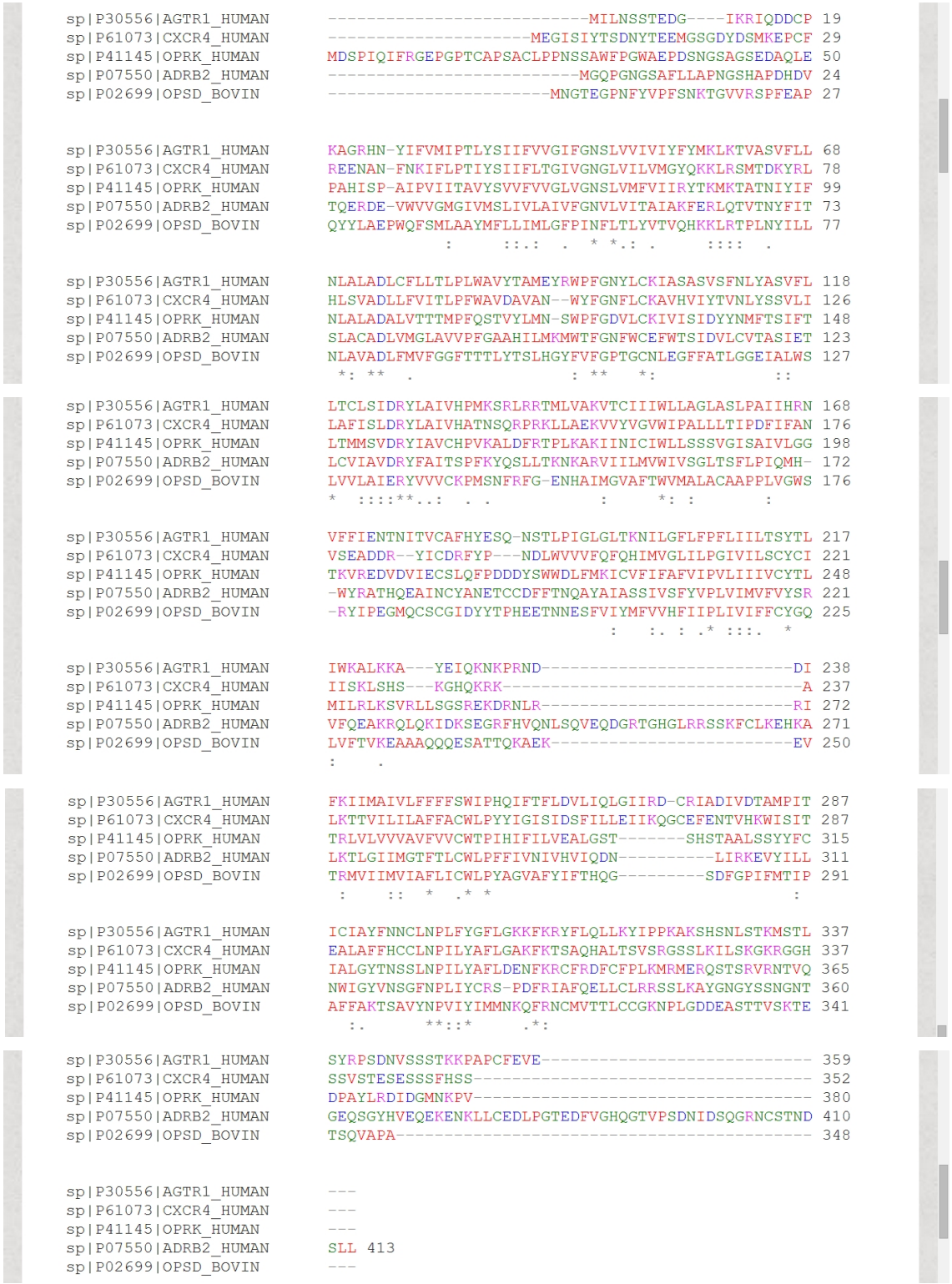


Fig. 5. Multiple sequence alignment between the target receptor AT1 and the template receptors, performed in CLUSTAL 2.1 [55].

The Homology Modelling protocol will proceed to arrange the backbone of the target sequence according to that of the template, using the sequence alignment to decide where to position each residue. Therefore, the quality of the sequence alignment is of crucial importance. Misplaced indels (gaps representing insertions or deletions) will cause residues to be misplaced in space. Although the routine used does alignments automatically, careful inspection and manual adjustment may improve the quality of the alignment, and hence, of the homology model.

After the multiple alignments were done, it was decided to perform iterative modelling on the base of each template, and advanced modelling on the base of multiple structures of the same template. Thus, it was done four iterative modelling protocols, and three advanced modelling protocols (kappa-type opioid receptor has only one available X-ray structure; therefore it hasn't been involved in multiple template construction).

Five models were produced from the four templates in iterative modelling, and 5 models were produced from each group of available structures from three templates in advanced modelling, so in total 35 models. The next step was to select the "best" models.

Iterative modelling			**Advance modelling**		
1. On the base of CXCR4 Chemokine receptor template			*1. On the base of 3 structures of CXCR4 Chemokine receptor*		
Model name	DOPE score	GA341 score	Model name	DOPE score	GA341 score
Model.It.1.1	-40877.5	0.95395	Model.Adv.4.1	-41308.8	0.99993
Model.It.1.2	-40389.2	0.95677	Model.Adv.4.2	-40507.5	0.99986
Model.It.1.3	-41117.5	**0.98503**	**Model.Adv.4. 3**	**-41484.0**	**0.99999**
Model.It.1.4	**-41491.1**	0.94713	Model.Adv.4. 4	-40401.9	0.98810
Model.It.1.5	-41117.0	0.96780	Model.Adv.4. 5	-40988.4	0.99831
2. On the base of β_2 adrenergic receptor template			*2. On the base of 2 structures of β_2 adrenergic receptor*		
Model name	DOPE score	GA341 score	Model name	DOPE score	GA341 score
Model.It.3.1	**-38933.3**	**0.29955**	**Model.Adv.2.1**	**-42005.4**	0.31227
Model.It.3.2	-38603.6	0.20276	Model.Adv.2.2	-40786.3	0.40978
Model.It.3.3	-38722.4	0.26280	Model.Adv.2.3	-41166.9	**0.44090**
Model.It.3.4	-38693.0	0.26446	Model.Adv.2.4	-40365.5	0.30343
Model.It.3.5	-38068.8	0.17337	Model.Adv.2.5	-41142.7	0.36918
3. On the base of Bovine Rhodopsine receptor template			*3. On the base of 5 structures of Bovine Rhodopsin receptor*		
Model name	DOPE score	GA341 score	Model name	DOPE score	GA341 score
Model.It.4.1	**-43537.9**	0.05072	Model.Adv.3.1	-42766.7	0.11459
Model.It.4.2	-43143.4	0.12643	**Model.Adv.3.2**	**-43578.2**	**0.12359**
Model.It.4.3	-42829.2	0.12070	Model.Adv.3.3	-43066.8	0.07247
Model.It.4.4	-42785.0	**0.18405**	Model.Adv.3.4	-43180.9	0.07225
Model.It.4.5	-43119.8	0.13055	Model.Adv.3.5	-42507.0	0.09612

4. On the base of K-type opioid receptor template		
Model name	DOPE score	GA341 score
Model.It.2.1	-37661.7	0.88938
Model.It.2.2	**-37896.0**	**0.93322**
Model.It.2.3	-37236.1	0.61869
Model.It.2.4	-37626.6	0.83868

Table 4. The summary of the successfully produced models.

After the assessment by DOPE and GA341 scores was done, and 7 best models were chosen, they were further assessed by the quality of secondary structure by calculating Ramachandran plot [57] for the whole protein. A Ramachandran plot is a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure. The quality of secondary structure elements was tested by using the definitions introduced by Morris et al. for the allowed and disallowed regions for φ and ψ angles of the backbone that reflects the stability of protein structure [58]. Basically, this analysis helps to say which protein model is better. The more amino acid residues fall into allowed region, the better the model. After Ramachandran-plot analysis of the chosen models, there were chosen 2 models: one from iterative modelling on the base of one chemokine receptor template – model 1, and one from advanced modelling on the base of three chemokine receptor templates – model 2. Their plots are presented on the Fig. 6. Model 1 has 93% of the residues in the allowed region, while model 2 has 94%.

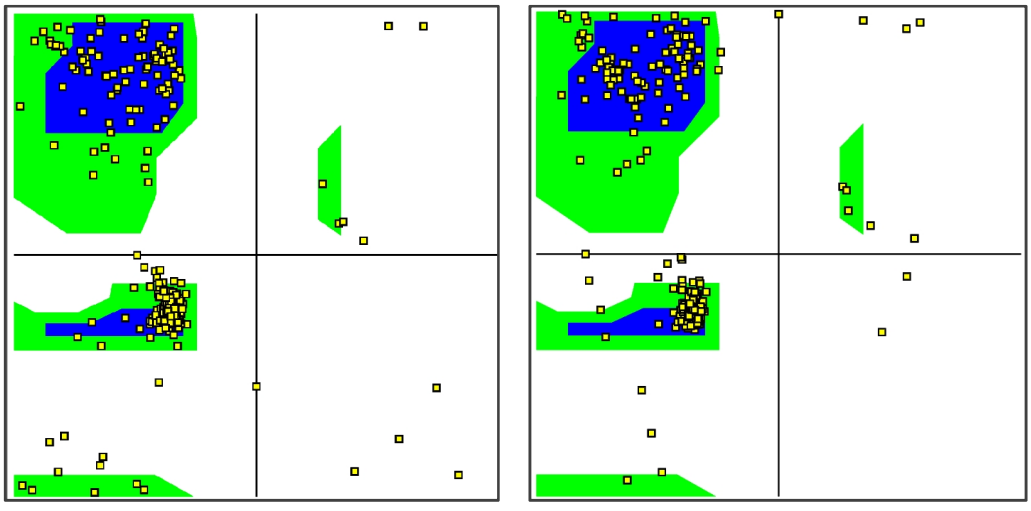


Fig 6. – Ramachandran plot: a) model 1; b) model 2.

Force-field minimization in AMBER: After two models were chosen, they were optimised with molecular mechanical level of theory in the Amber software [59]. Each of the modelled receptors was placed to the box, the faces of which are 12 Å distant from the

protein, and the boxed was filled with explicit water molecules (TIP3PBOX) with *xleap* program (Fig. 7). Chlorine or sodium counter ions were added to keep the system neutral. The relaxation was carried out in three steps. In a first phase only the geometry and position of the water molecules were minimized by keeping the rest of the system fixed. This step aims to reduce possible unfavourable interactions caused by automated placement of water molecules by *xleap* program. In a second step the position of the hydrogen atoms was optimized, while maintaining the other atoms fixed. At third step the position of all atoms of the model was optimized. The Force Field used was parm99SB, specifically created for proteins.

The disulfide bonds were formed between Cys101 in TM3 and Cys180 in extracellular-loop 2 by covalently bound their sulphur atoms (Fig. 7). Globally, we have performed energy minimizations with the AMBER force field to allow the protein atoms to relax and adjust.

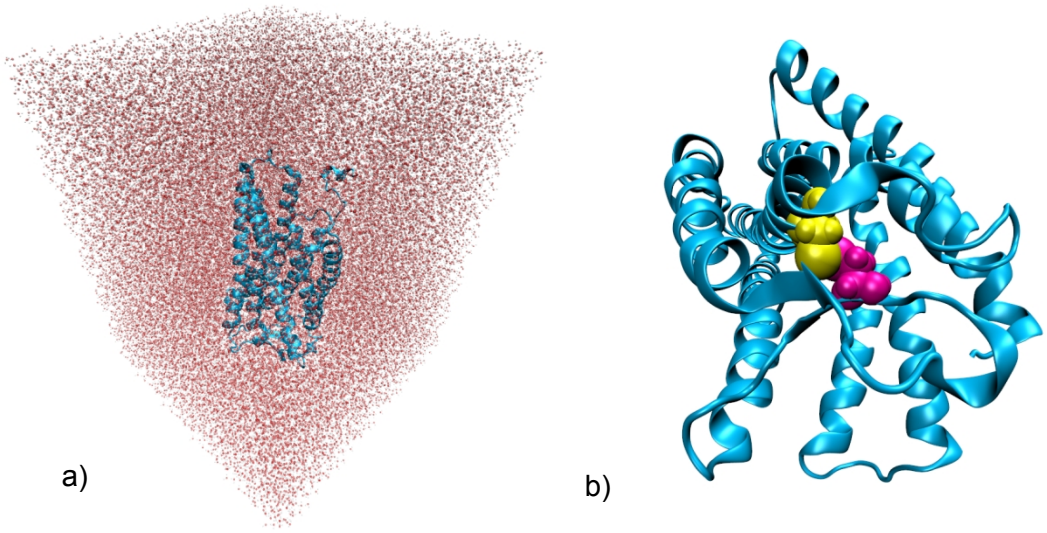


Fig. 7 – Force-field minimization in AMBER: a) AT1 receptor in the water box; b) modelled AT1 receptor with the conserved disulphide bond Cys101-Cys180 (magenta and yellow).

After energy minimization was done, the models were visually checked. The following residues experimentally shown to be involved in Ang-II binding, Lys199, Asn111, His256, Gln257, Asn294, Asn295 and Asn298, are indeed pointing inward. The analysis of interactions would provide a clear understanding which model would be preferable in further studying after molecular docking calculations would be performed.

Now, the models are ready to proceed with the next step of study – Molecular Docking.

4. MOLECULAR DOCKING

4.1 Protein-Ligand Docking methodology

Molecular docking is a widely used computer simulation procedure to predict the structure of the intermolecular complex formed between two or more constituent molecules, and tries to distinguish, from the energy point of view, binding mode of such complex, i.e. pose that maximizes the binding energy of interaction while minimizing the total energy of the complex. The accurate prediction of the binding poses between the ligand and protein is of fundamental importance in modern structure-based drug design.

The present study is related to the protein-ligand docking. The protein-ligand docking tries to predict the position and orientation of a ligand (a small molecule) when it is bound to a protein, enzyme or *receptor* (macromolecule). In our study we docked the natural peptide Ang-II and known inhibitors to a modelled receptor AT1 (Fig. 9).

Modelling the interaction of two molecules is a complex and challenging problem. Many forces are involved in the intermolecular association, including hydrophobic, van der Waals, or stacking interactions between aromatic amino acids, hydrogen bonding, and electrostatic forces. Modelling the intermolecular interactions in a ligand-protein complex is difficult since there are many degrees of freedom as well as insufficient knowledge of the effect of solvent on the binding association. The process of docking a ligand to a binding site tries to mimic the natural course of interaction of the ligand and its receptor via the lowest energy pathway. There are simple methods for docking rigid ligands with rigid receptors, and flexible ligands with rigid receptors. However, methods of docking that considers conformationally flexible ligands and receptors are problematic, as well as, large time and computationally consuming. Therefore, in our study an approach to consider the ligand flexibility while assuming having a rigid protein receptor was used.

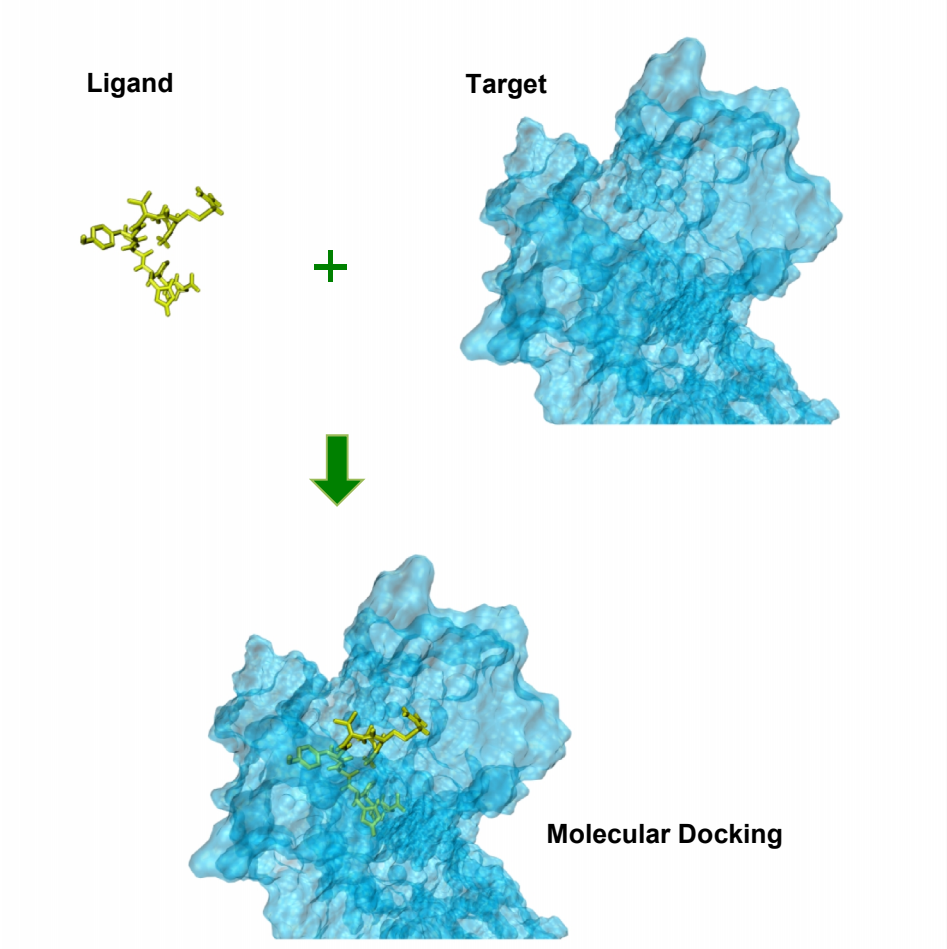


Fig. 8 – Representation of the Protein-Ligand Molecular Docking

Docking protocols can be described as a combination of a search algorithm, and the scoring functions. The search algorithm should create an optimum number of configurations that include the experimentally determined binding modes. Although a rigorous searching algorithm would go through all possible binding modes between the two molecules, this search would be impractical due to the size of the search space and amount of time it might take to complete. As a consequence, only a small amount of the total conformational space can be sampled, so a balance must be reached between the computational expense and the amount of the search space examined. On the other hand, scoring function consists of a number of mathematical methods used to predict the strength of the non-covalent interaction called the binding affinity. In all the computational methodologies, one important problem is the development of an energy scoring function that can rapidly and accurately describe the interaction between the protein and ligand. Over the course of the last years, different scoring functions have been developed that exhibit different accuracies and computational efficiencies. Some of these commonly-used functions are: force-field, empirical, knowledge-based and consensus scoring.

The molecular docking study was performed using software AutoDock v.4.0 [60]. This software contains a package of automated tools that were designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor with given 3D structure.

AutoDock consists of two main programs: AutoDock that performs the docking of the ligand to a set of grids describing the target protein, and AutoGrid that pre-calculates these grids. In order to optimise the creation of the input files that were required to run AutoDock and AutoGrid, the VsLab plug-in [61], which was written in the group of Theoretical Chemistry and Computational Biochemistry of the University of Porto, was used. VsLab is an easy-to-use and freely available graphical interface for the molecular docking software AutoDock that has been included into VMD as a plug-in.

AutoDock offers a variety of search algorithms to explore a given docking problem. These include Monte Carlo Simulated Annealing (SA); a Genetic Algorithm (GA); and a hybrid local search GA, also known as the Lamarckian Genetic Algorithm (LGA). In general, the LGA performs the best out of SA, GA, and LGA, in finding the lowest energy of the system. We chose LGA algorithm for our studies.

4.2 Docking of the angiotensin II to the model of the AT1 receptor

After two models of the AT1 receptor were prepared for the study of docking, Ang-II was docked to both models following the same procedure. The octapeptide Ang-II was docked to the docking area that includes the binding pocket according to known site-directed mutagenesis data [62, 63, 64, 65]. The parameters of the docked box, used in the study on the base of known interactions, were as following: width 25.87 Å; height: 26.625 Å; depth: 31.5 Å. Number of grid points at X, Y, Z: 66, 76, 89, with grid spacing of 0.375 Å (Fig. 9).

The ligand, peptide hormone, Ang-II was obtained from the PDB (PDB code 1N9V).

As can be seen from the representation of the docked Ang-II to the modelled AT1 receptor (Fig. 9), it is nestled between TM helices.

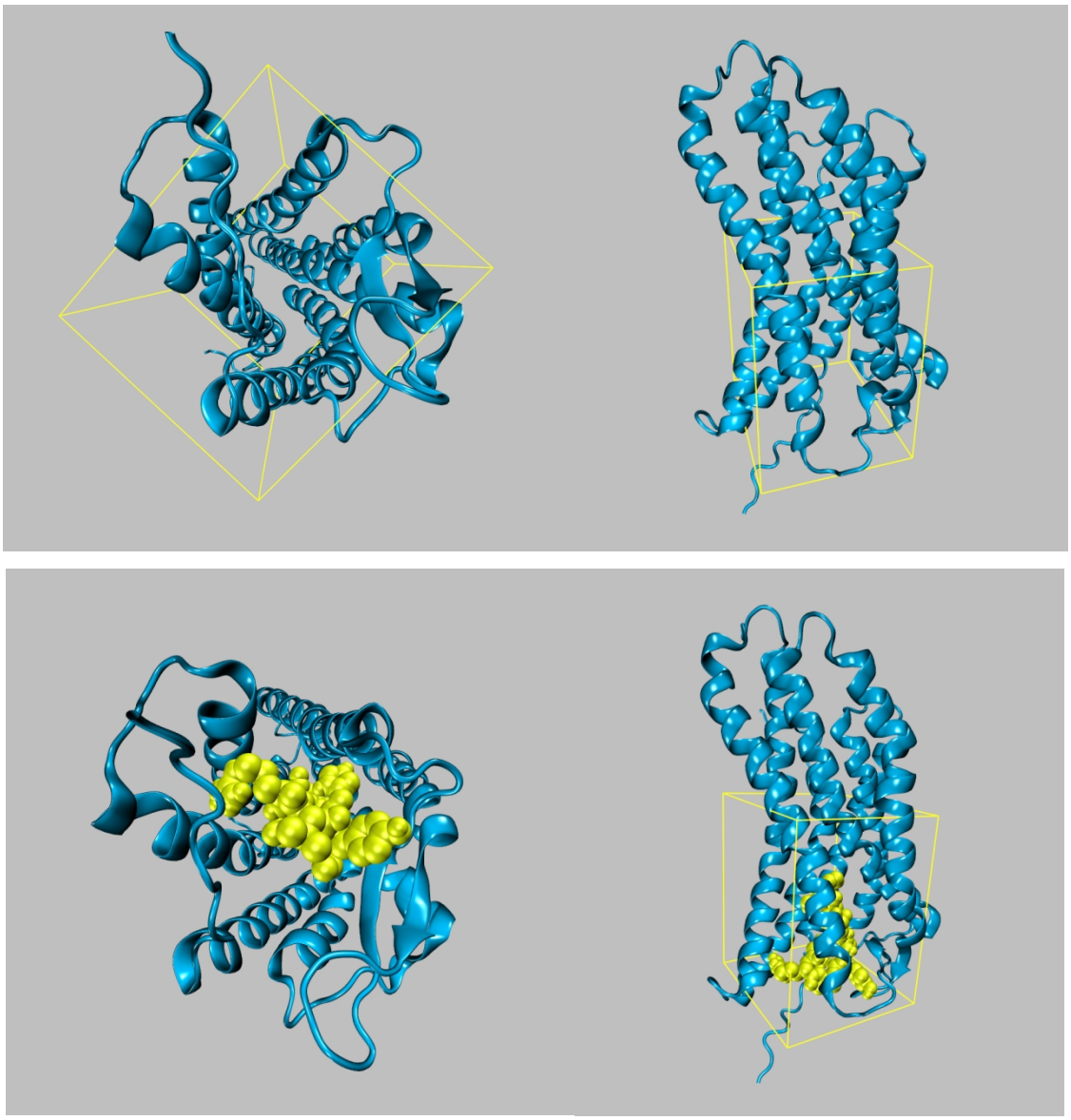


Fig. 9 – Representation with the box of the docking area of the AT1 receptor (blue), and docked Ang-II (yellow).

All the docking poses of Ang-II to the AT1 receptor were analysed for binding site to choose the best model. The results of docking showed better results for the second model (on the base of multiple structures of the chemokine receptor). The discussions of docking are presented below. This model was used in further analysis – docking of the inhibitors.

Table 5 contains solutions of the docking of the ligand with corresponded energy score that is the Gibbs energy of interaction. The lower the value the stronger ligand is bound to the receptor. The dissociation constants were calculated as well.

Ligand (S#, docking solution number)	Score	Ki(M)
AngII-S1	-10.46	$2.15 \cdot 10^{-08}$
AngII-S2	-8.79	$3.60 \cdot 10^{-07}$
AngII-S3	-8.40	$6.95 \cdot 10^{-07}$
AngII-S4	-8.19	$9.91 \cdot 10^{-07}$
AngII-S5	-8.18	$1.01 \cdot 10^{-06}$
AngII-S6	-6.60	$1.45 \cdot 10^{-05}$
AngII-S7	-6.58	$1.50 \cdot 10^{-05}$
AngII-S8	-6.47	$1.81 \cdot 10^{-05}$
AngII-S9	-6.40	$2.03 \cdot 10^{-05}$
AngII-S10	-5.34	$1.22 \cdot 10^{-04}$

Table 5 – Solutions of the docking of Ang-II to the model of the AT1 receptor, with corresponding energy scores, and dissociation constants.

The best docked solutions were visualised thoroughly with VMD, whether they fulfilled known experimental data (mainly the interaction of the carboxyl terminal with Lys199, and Arg167). The first complex, which showed the better and most stable binding free energy value, was chosen for further investigation, and the interactions of which are described.

As presented on Fig. 10, the representative binding conformation of Ang-II shows a hydrogen bond interaction between Lys199 in TM5 and the carboxyl terminal of the peptide (Phe8). The distance between oxygen of the carbonyl terminal of the peptide and NH_3 group of Lys199 of the AT1 receptor are 2.64 Å and 2.88 Å. Another important interaction was found between the carboxyl terminal of the peptide (Phe8) and Arg167 in TM4 of the AT1 receptor. This interaction is critical, since studies have showed an almost complete loss of Ang-II affinity when Arg167 was mutated [66, 67]. In the present homology model, Arg167 points to the binding site core (Fig. 10). And the distance between oxygen of the carboxyl terminal of the peptide (Phe8) and NH_2 group of Arg167 of the AT1 receptor are 3.32 Å and 4.18 Å. The interaction of His256 with Phe8 (Fig. 10) known to be critical for Ang-II activation is a stacked attraction of aromatic rings [68, 69, 70]. In the present homology model the rings are oriented parallel to each other with the distance approximately of 6.00 Å.

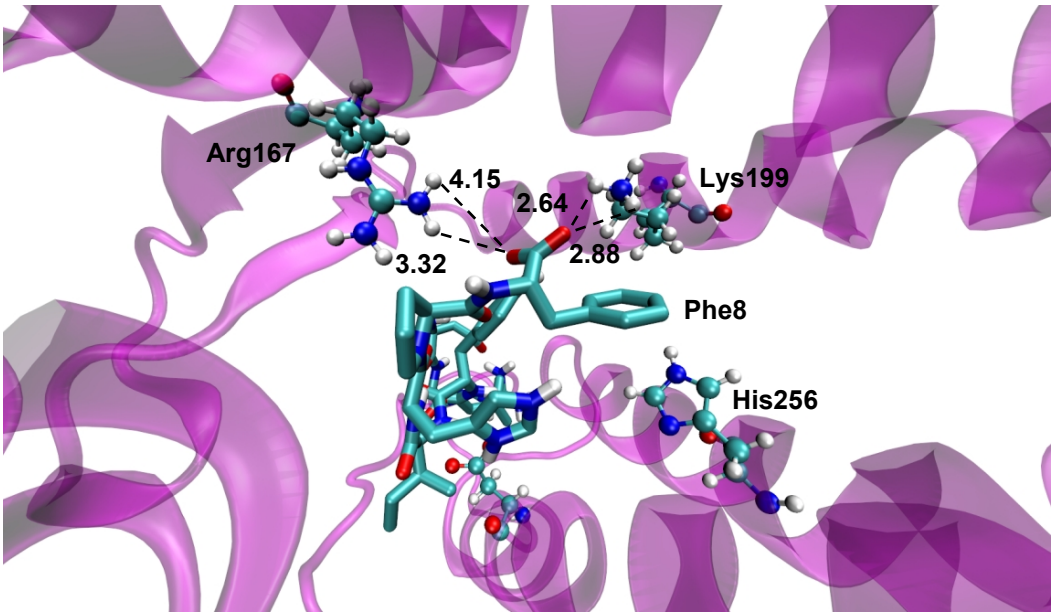


Fig. 10 – Representation of the predicted binding site of Ang-II (licorice) in the AT1 receptor (NewCartoon magenta). Residues Lys199, Arg167, and His256 of AT1 receptor are represented in CPK. The distance unit is specified in angstroms.

The Phe8 bulky ring of Ang-II is oriented towards the highly hydrophobic pocket formed by residues Ala104, Val108, Phe77, Leu81, Trp253, Leu112, Tyr292, Phe182, Trp84, Met284, Ile288 of the AT1 receptor (Fig. 11).

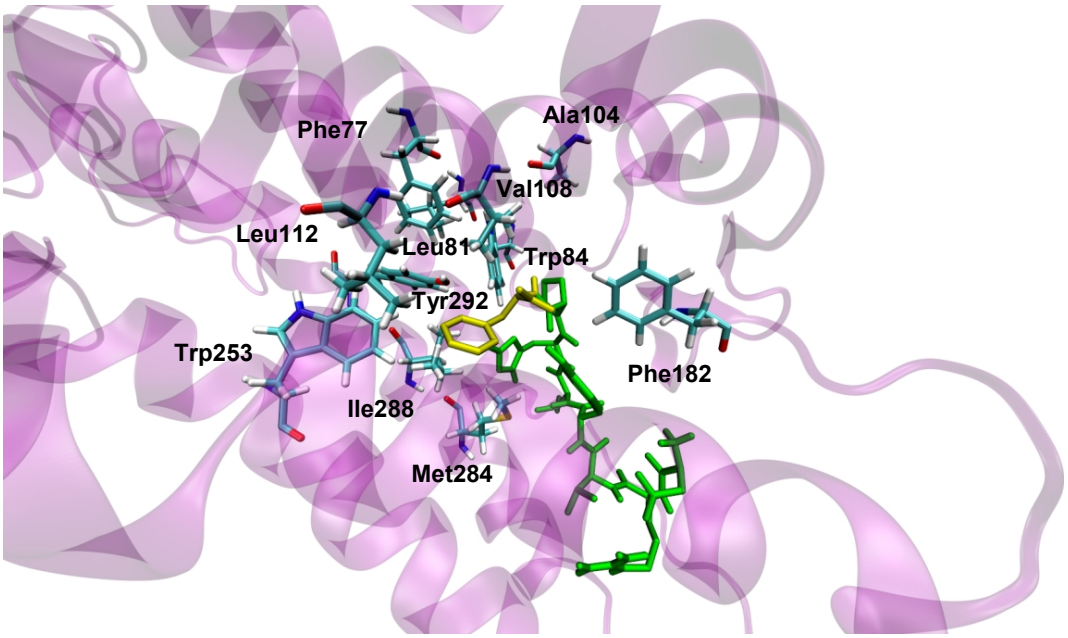


Fig. 11 – Representation of the Phe8 of Ang-II (yellow) in the hydrophobic pocket formed by residues Ala104, Val108, Phe77, Leu81, Trp253, Leu112, Tyr292, Phe182, Trp84, Met284, Ile288 of the AT1 receptor.

Arg2 interacts with the charged side chain of Asp263 (Fig. 12). Asp263 mutation has also been reported to almost abolish binding of the antagonistic peptide [Sar,Ile]Ang-II

to the AT1 receptor [71]. The distances between oxygen of Asp263 and NH_2 group of Arg2 are 1.83 Å and 3.23 Å. Asp281, which have been documented as important for full agonism of Ang-II, interacts with Tyr4 and Ile5 of Ang-II. The suggested orientation of Tyr4 and Ile5 towards TM7, by forming polar interactions with D281, has been also reported recently [72].

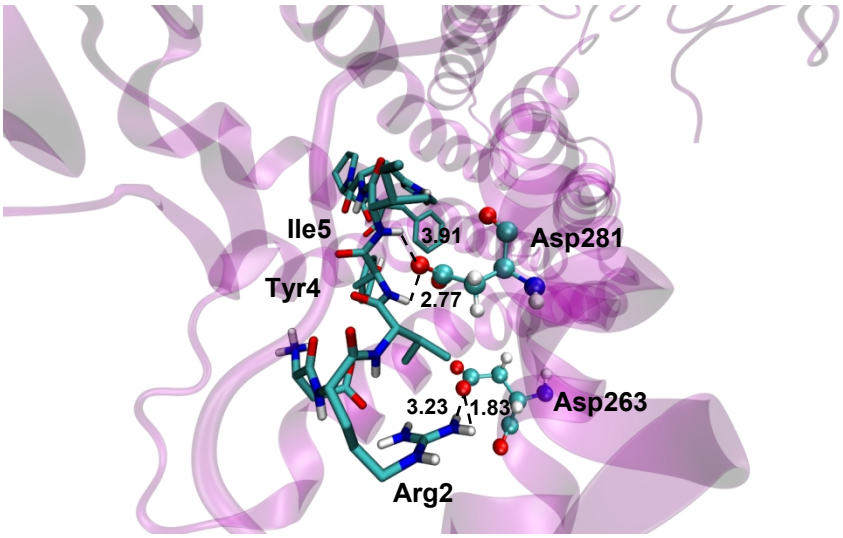


Fig. 12 – Representation of the interaction between Tyr4 and Ile5 of Ang-II and Asp281 of the AT1 receptor, and the interaction between Arg2 of Ang-II and Asp263 of the AT1 receptor. The distance unit is specified in angstroms.

4.3 Docking of the known inhibitors to the model of the AT1 receptor

Each of non-peptide inhibitors (Table 1) was docked to the same docking area as was used for Ang-II. The geometries of non-peptide ligands were obtained from ChemSpider database and the DrugBank database [73, 74].

The predicted binding pose of the inhibitor *losartan* matches very well with available site-directed mutagenesis data, and is presented on the Fig. 13. Lys199 creates a weak hydrogen bond interaction with the losartan hydroxyl group (4.01 Å), while the tetrazole ring is cradled between Tyr184 and Leu195 (3.91 Å and 2.22 Å, respectively). The central phenyl rings are stabilized by pi-stacking, one with Tyr113, and another with Phe182. Imidazole ring of losartan is stabilized with hydrogen bond with Tyr292 (3.32 Å). The important residue His256 also interacts with a hydroxyl group (2.03 Å).

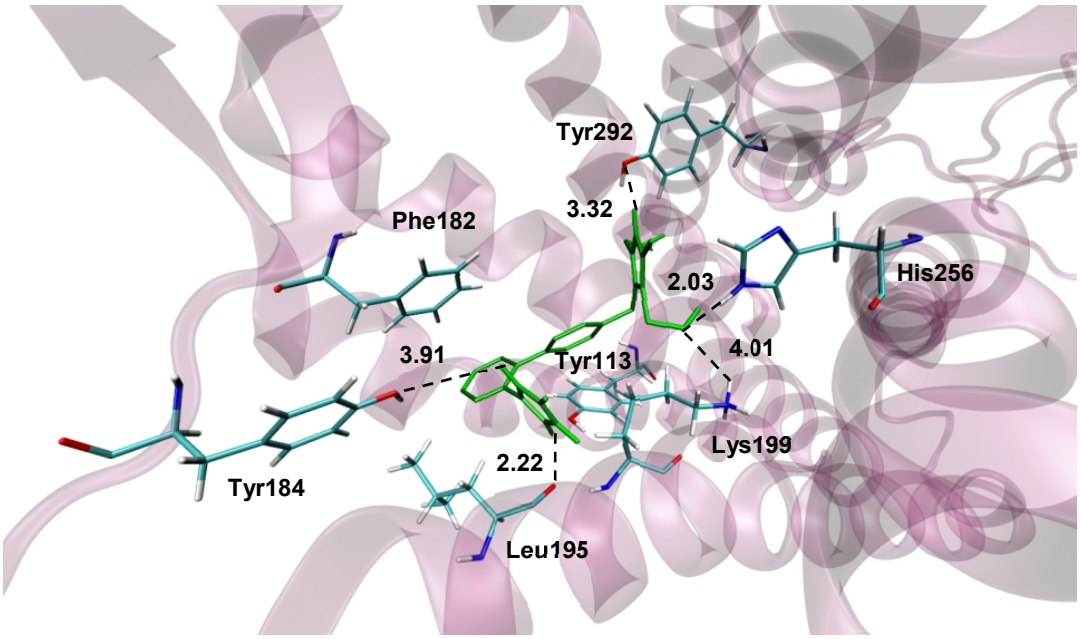


Fig. 13 – Representation of the predicted binding site of losartan in the AT1 receptor. The distance unit is specified in angstroms.

The predicted binding site for *cardesartan* (Fig. 14) is similar to that one of losartan. The carboxyl group of the inhibitor creates hydrogen bonds with residues Lys199 and His256 (4.28 Å and 1.72 Å, respectively). Tetrazole ring of the inhibitor is stabilized by

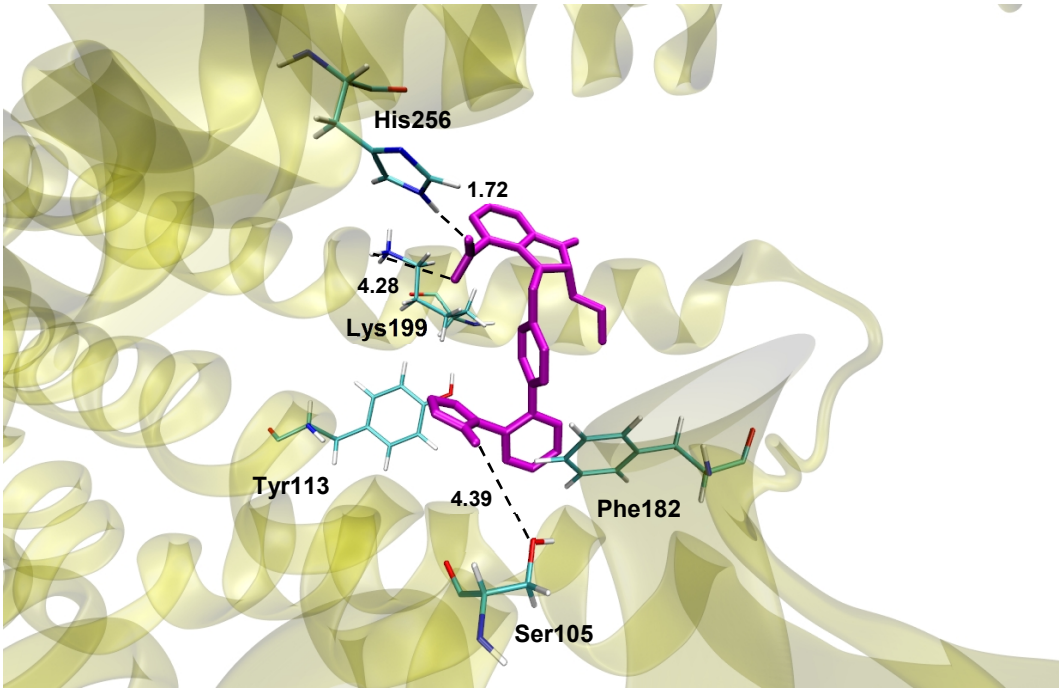


Fig. 14 – Representation of predicted binding site of candesartan in the AT1 receptor. The distance unit is specified in angstroms.

pi-stacking with Tyr113 (the rings are on the distance around 5.0 Å), and by the interaction with hydroxyl group of Ser105 (4.39 Å). Central ring is stabilized by pi-stacking with Phe182.

The representative binding conformation of *olmesartan* is depicted on Fig. 15. The predicted binding pocket is similar to the previous non-peptide inhibitors, but in this case Lys199 is not in good orientation to the carboxyl group, but still quite close on the distance of 3.90 Å to it. The same carboxyl group, again as in previous inhibitors, is stabilized by carboxyl group of His256 (2.51 Å). Tetrazole cycle is stabilized by pi-stacking with Tyr113 (the distance is around 5.50 Å), and by interaction with Ser105 (4.30 Å), while phenyl cycle is stabilized by pi-stacking with Phe182 (distance around 4.50 Å).

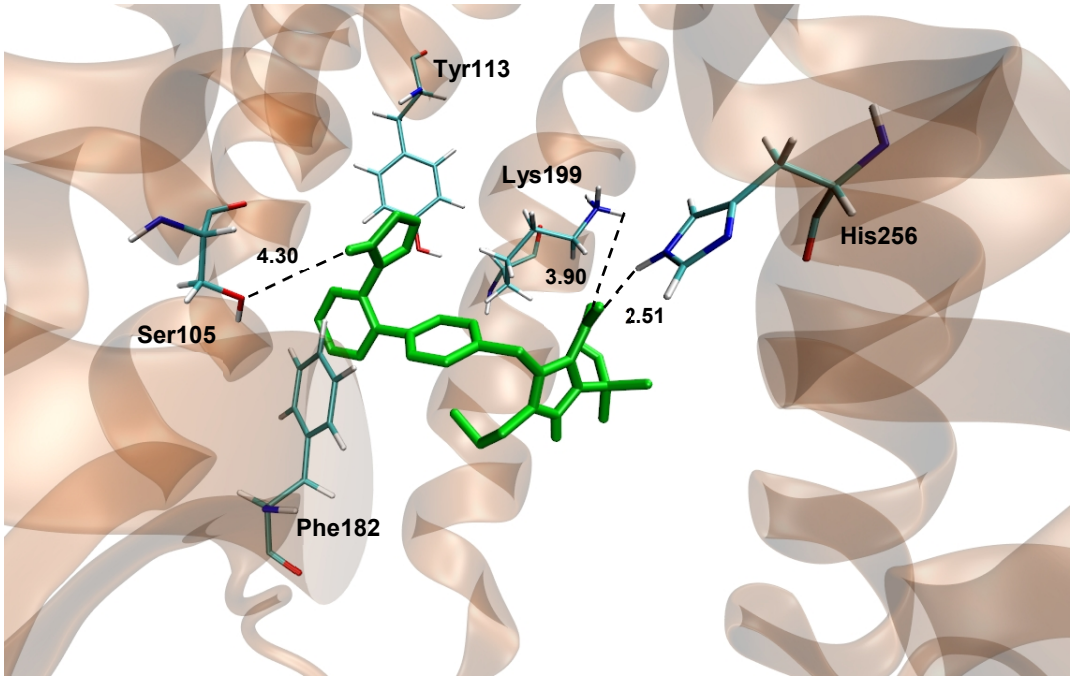


Fig. 15 – Representation of predicted binding site of olmesartan in the AT1 receptor. The distance unit is specified in angstroms.

The predicted binding site for *eprosartan* is in a good agreement with previous results. Lys199 is in good orientation and close to the carboxyl group creating an interaction with it (1.82 Å). Ser105 stabilizes the imidazole ring (4.88 Å), while His256 stabilizes carboxyl groups (3.17 Å and 2.03 Å). Phenyl cycle is stabilized by pi-stacking with Trp253 (4.50 Å), while imidazole is pi-stacked with Phe182 (4.00 Å).

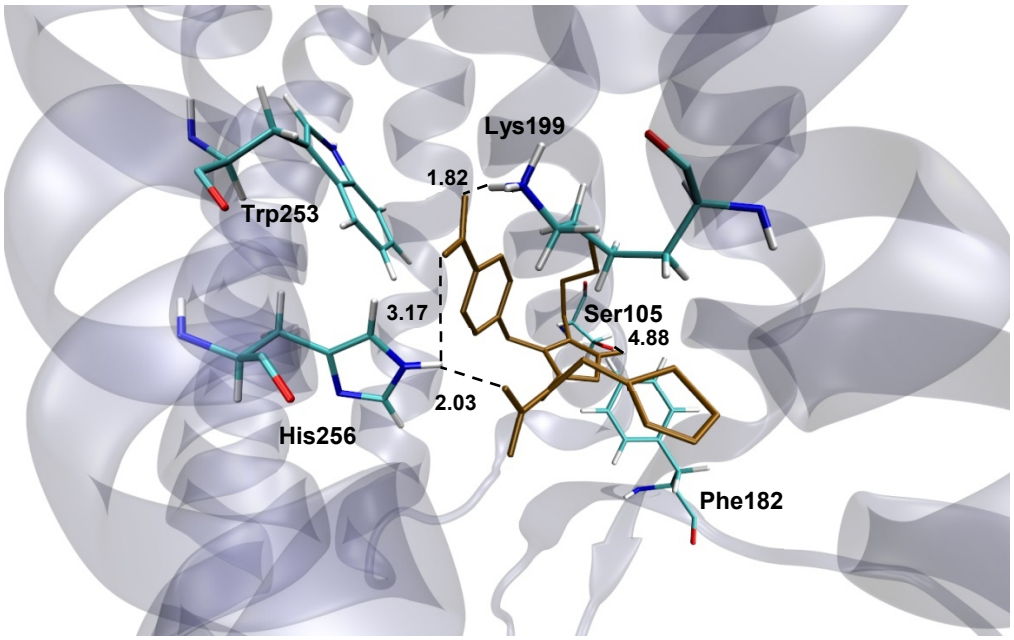


Fig. 16 – Representation of predicted binding site of eprosartan in the AT1 receptor. The distance unit is specified in angstroms.

The predicted binding pose of the inhibitor *irbesartan* (Fig. 17) includes interaction between hydroxyl group of Tyr184 and tetrazole (2.07 Å and 2.08 Å). Imidazole ring is stabilized by hydroxyl groups of Tyr113 and Ser109 (2.23 Å and 2.67 Å). Lys199 is close but as in the case of olmesartan is not in good geometrical orientation. Phenyl cycle is stabilized by pi-stacking with Phe182 (6 Å).

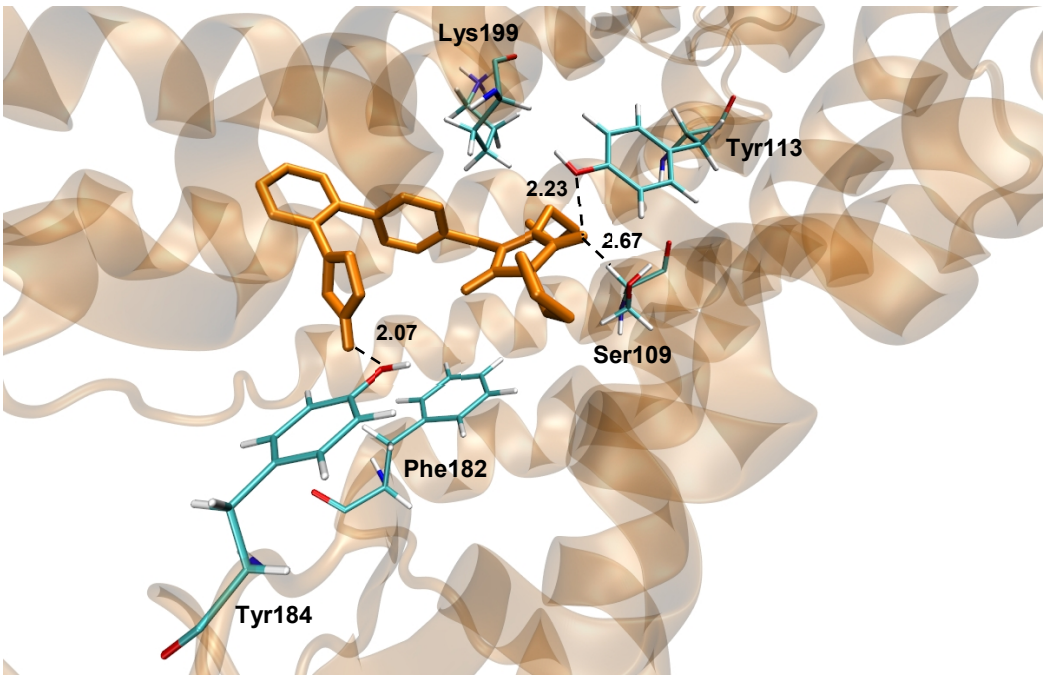


Fig. 17 – Representation of predicted binding site of irbesartan in the AT1 receptor. The distance unit is specified in angstroms.

Docking of the inhibitor *telmisartan* revealed the binding site similar to the previous inhibitors (Fig. 18). The carboxyl group of the inhibitor has interactions with Lys199 and His256 (5.72 Å and 2.39 Å, respectively), while imidazole cycle has interactions with hydroxyl groups of Tyr113 (2.05 Å) and Ser109 (3.16 Å). Stabilization of phenyl cycle again occurs through pi-packing with Phe182.

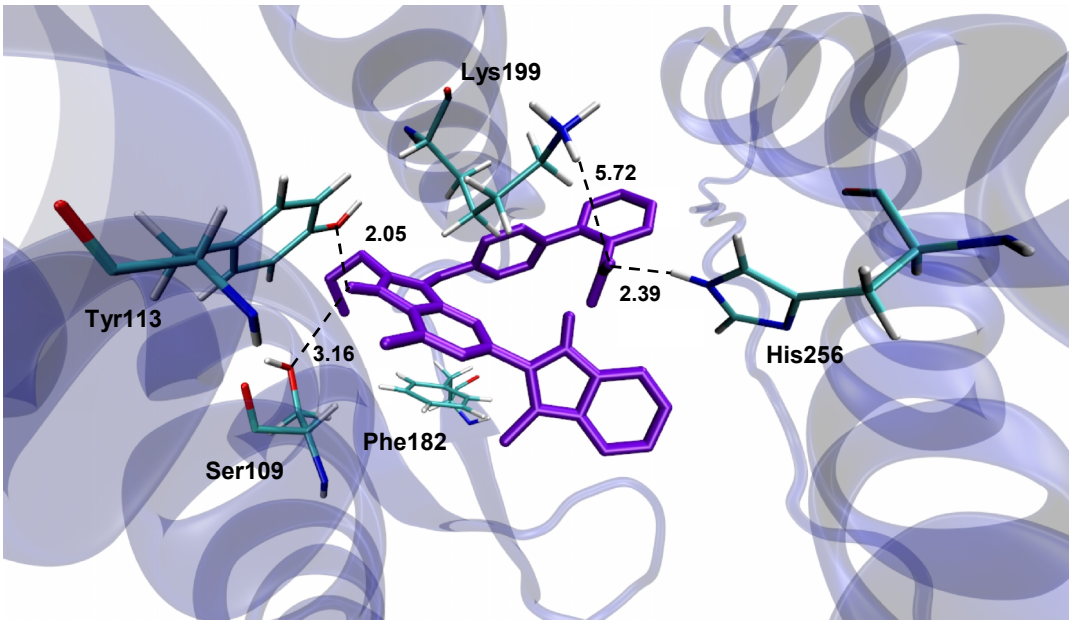


Fig. 18 – Representation of predicted binding site of telmisartan in the AT1 receptor. The distance unit is specified in angstroms.

The representative binding conformation of *valsartan* is shown in Fig. 19. Carboxyl group of the inhibitor interacts with Lys199 (3.37 Å) and His256 (2.78 Å). Imidazole cycle

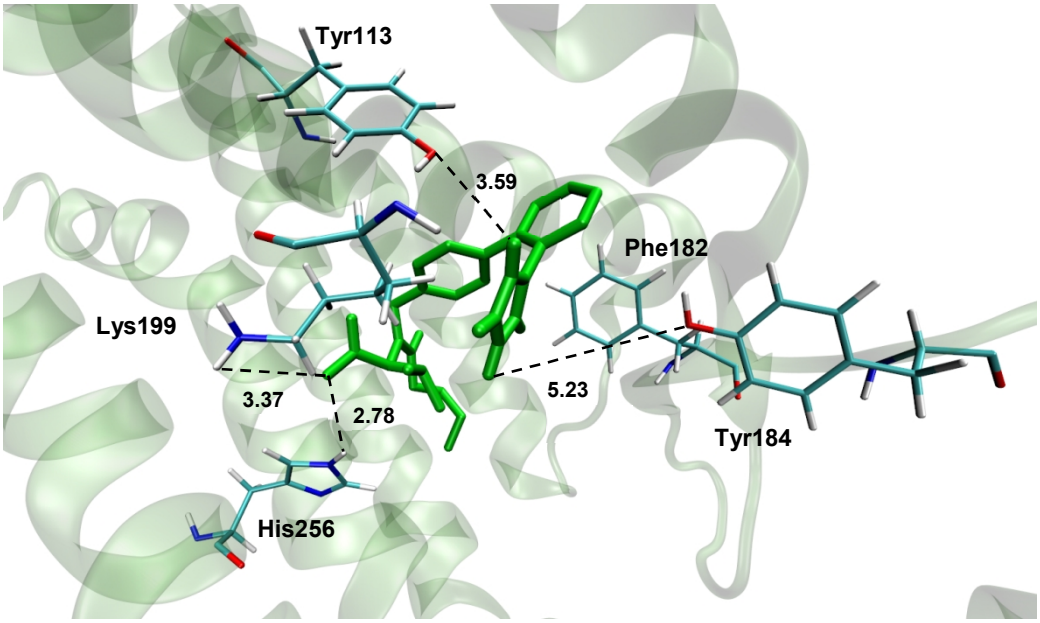


Fig. 19 – Representation of predicted binding site of valsartan in the AT1 receptor. The distance unit is specified in angstroms.

is in between hydroxyl groups of Tyr113 (3.59 Å) and Tyr184 (5.23 Å). Phenyl cycle is pi-stacked with Phe182 (5.00 Å).

The predicted binding pose of the last inhibitor *azilsartan* has a similar pose as the inhibitors described above. The carboxyl group of azilsartan is cradled between Lys199 and His256 (3.30 Å and 3.67 Å, respectively), Asp263 creates a hydrogen bond interaction with hydrogen of imidazole heterocycle (2.29 Å), Tyr184 creates a hydrogen bond interaction with a terminal heterocycle (2.21 Å), the same ring is stabilized by pi-stacking with Phe182 (3.50 Å).

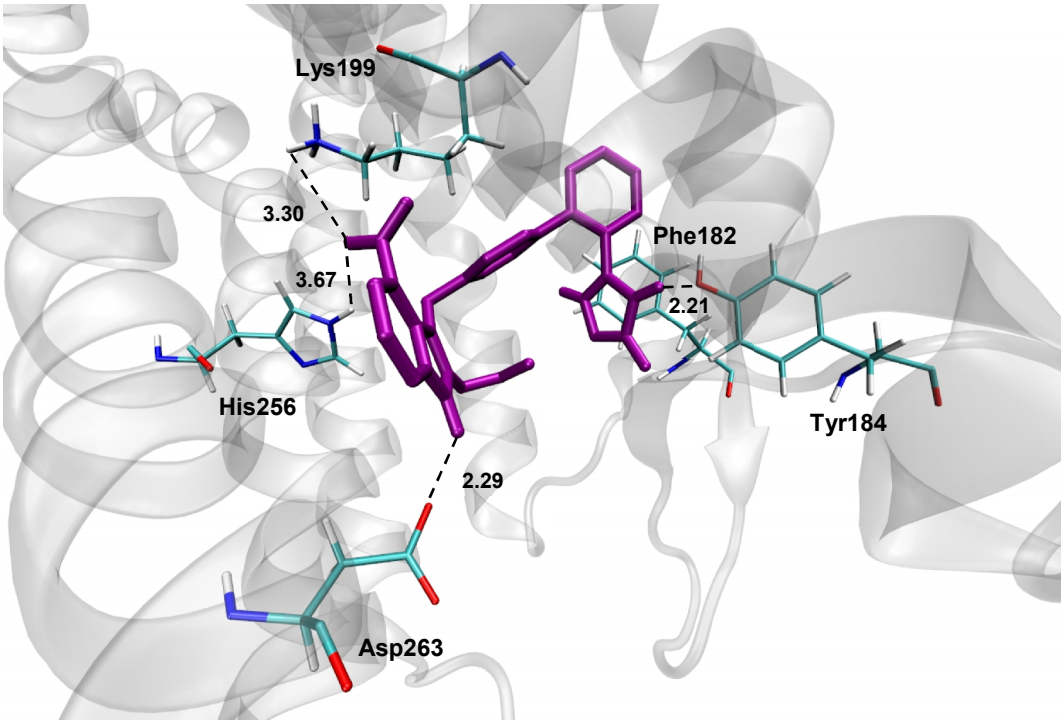


Fig. 20 – Representation of predicted binding site of azilsartan in the AT1 receptor. The distance unit is specified in angstroms.

According to the docking results, the important residues of the binding pocket that interacts with inhibitors are Lys199 and His256, creating hydrogen bonds with carboxyl groups, Phe182 that stabilizes phenyl cycles by pi-stacking, Tyr113 and Ser109 in majority of the cases show interaction with the carboxyl group. Tyr184 is often interactions with inhibitors.

Below in the Table 6 are summarized Gibbs energies (ΔG) obtained from the docking calculations, as well as experimental values found in the PubMed database [75]. The inhibitors are ranked according to the energy score. The results are reasonably in accordance with the experiment.

№	Inhibitor	Calculated ΔG (kcal/mol)	Experimental ΔG (kcal/mol)
1	Telmisartan	-11.20	-11.50
2	Azilsartan	-9.02	-
3	Irbesartan	-8.63	-11.87
4	Eprosartan	-8.62	-
5	Olmesartan	-8.19	-13.70
6	Candesartan	-8.03	-11.6
7	Valsartan	-7.84	-10.4
8	Losartan	-7.80	-9.58

Table 6 – Ranking of the inhibitors by their Gibbs energy values.

The obtained results of the docking validate the docking protocol. And the same protocol can be used for further investigation.

Conclusions

In the current study insights for the design of novel drugs for hypertension were provided by applying computational chemistry approach. Homology Modelling was applied to create a model of the 3D structure of the AT1 receptor. The binding poses of Angiotensin II to this receptor, which is an actual cause of the high pressure in the organism, were determined using Molecular Docking. The docking protocol was validated by docking of several AT1 antagonists to the created model of the AT1 receptor. This protocol can be used in the future investigation of obtaining novel AT1 receptor blockers using Virtual Screening.

5. PROSPECTIVE FOR THE FUTURE

After the 3D structure of the target AT1 receptor was obtained through Homology Modelling technique and the protocol was validated with Molecular Docking, it is ready to be used in the next step of drug development – hit identification.

Hit identification – is a docking methodology, combined with a scoring function that can be used to quickly screen large databases of potential drugs *in silico* to identify molecules that could interact with a target, or have similar properties to known ligands, and become a promising hit or lead compound. Afterwards these presumed molecules can be synthesized, tested and optimized in the laboratory. These types of methodologies are commonly known as Virtual High Throughput Screening or simply Virtual Screening (VS). VS is revealed as an important tool in drug development of novel drugs. There is a wide range of methodological protocols available in screening databases for

the lead compounds. The number of methods and software packages which employ the target and ligand based virtual screening are increasing rapidly, making of high interest the applicability of various virtual screening methods in the new lead identification and optimization.

In the current project, afterwards the databases will be screened against the AT1 model in order to find lead compounds. There are many available databases in the net, free of charge that can be used. One of particular interest is ZINC database [76] that contains over 13 million purchasable compounds for virtual screening. Subsequently, the hits obtained from the virtual screening have to be docked to the receptor model, and best-scored structures of the system receptor – ligand would be subjected to Molecular Dynamics simulations. Having identified the best candidates, they could be proceeding with lead optimization. The information obtained from the study may in turn be used to design more potent and selective analogues. Appropriate size and electrophilicity substituents will be placed to complement the morphology and electrostatics of the binding pockets.

Virtual Screening often requires careful preparation of both target and compound library, use of optimal parameters as well as careful analysis of the results. That is why the careful testing of the model of the AT1 receptor was done.

In the further study, the validated docking protocol would be applied to the ZINC database [76]. To screen for a compounds that can bind to the created model of the AT1 receptor, and in future can be a potential lead compounds for hypertension.

II. THE MODELLING OF THE ENZYME PATGMAC SYSTEM

1. INTRODUCTION

In the second part of the work and in order to learn other computational methodologies, there was performed the preparation of the enzyme macrocyclase PatG (PatGmac) system for the study of its catalytic mechanism. This will be useful to characterise and potentially adapt their functions for use in the production of novel cyclic peptides. This enzyme catalyses the macrocyclization of linear peptides designated "patellamides" [77, 78].

The largely unexplored world of marine natural products has attracted the attention of biologists and chemists over the past decades. Marine organisms elaborate pharmaceutically useful compounds, therefore are a recognized source of diverse biologically active molecules.

Patellamides, members of the cyanobactin family, were isolated from the ascidian *Lissoclinum Patella* [79, 80, 81]. *L. patella* harbours an obligate symbiont, *Prochloron didemni*, which has the necessary genes to produce patellamides, cyclic peptides that display a versatility of biological activities such as cytotoxicity, and the ability to reverse multiple-drug resistance in human leukemia cells [82, 83]. This makes these compounds very attractive in becoming potential drugs in cancer treatment. But to date their production has proven challenging for synthetic chemists.

		R_1	R_2	R_3	R_4	Sequence
1	Pat A	CHMeEt	$CHMe_2$	CHMeEt	$CHMe_2$	ITVCISVC
2	Pat B	CH_2CHMe_2	CHMeEt	CHMeEt	CH_2Ph	LTICITFC
3	Pat C	$CHMe_2$	Me	CHMeEt	CH_2Ph	VTACITFC
4	Pat D	CHMeEt	Me	CHMeEt	CH_2Ph	ITACITFC
5	Pat E	$CHMe_2$	$CHMe_2$	CHMeEt	CH_2Ph	VTVCITFC
6	Pat F	$CHMe_2$	$CHMe_2$	$CHMe_2$	CH_2Ph	VTVCVTFC
7	Pat G	CHMeEt	Me	CH_2CHMe_2	CH_2Ph	ITACLTFC

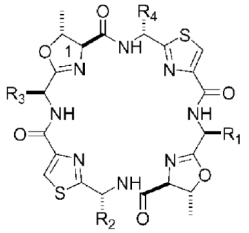


Fig. 21 – Structures of the known patellamides isolated from *L. patella* and their peptide sequences [83].

The patellamides are cyclic octapeptides containing highly conserved thiazole and oxazoline and D-stereo centres (Fig. 21). The patellamides are derived from a precursor peptide, PatE, which undergoes several modifications, catalysed by the enzymes PatA, B,

C, D, F and G, to form the final cyclic peptide (macrocycle). These modifications include heterocyclisation, proteolysis, oxidation and macrocyclisation.

2. BIOSYNTHESIS OF PATELLAMIDES

The biosynthesis of such patellamides is complicated, with many steps. In brief, the patellamides are produced by a microcin-like biosynthetic pathway in which the precursor peptide gene (patE) directly encodes the amino acid sequence of the patellamide product(s). The post-translational modification machinery encoded by the other genes in the cluster, then act on this precursor peptide to give the fully processed cyclic peptide(s). The leader sequence of the precursor peptide PatE was assumed to target it toward the post-translational machinery (scheme on the Fig. 22). The cyclic-peptide sequences within the precursor peptide are flanked by start/stop recognition motifs that recruit enzymes for post-translational modifications. PatGmac has been reported to recognize the C-terminal macrocyclization signature AYDG [84]. The sequence of the biosynthetic events suggested by Schmidt et al. [80] starts with the heterocyclisation by PatD2 to form the oxazoline and thiazoline rings. Thiazoline is then oxidised to thiazole by PatG1. Cleavage of the mature PatE could occur under the influence of PatG2 or PatA, followed by adenylation by PatD1 and macrocyclisation, which could be spontaneous or enzymatically controlled.

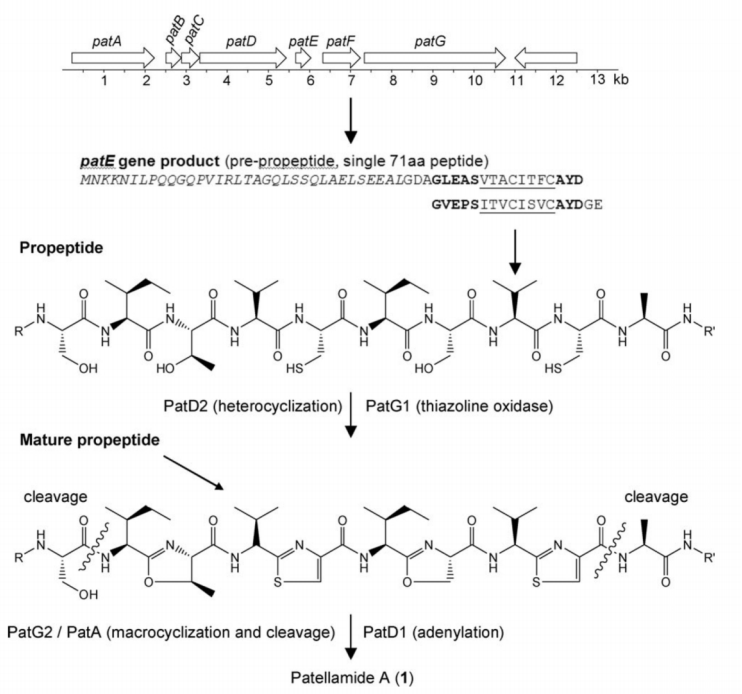


Fig. 22 – Scheme of suggested biosynthetic pathway for patellamide [81].

3. PROPOSED MECHANISM FOR MACROCYCLIZATION

In the current project we aim to prepare the system for studying the step of macrocyclisation. The macrocyclization is carried out by the catalytic triad Asp548-His618-Ser783. The proposed mechanism (Fig. 23) consider nucleophilic attack of the serine of the active site to the carbon of carbonyl with forming enzyme-substrate intermediate. Subsequently, the four stop recognition residues AYDG are cleaved and N terminus of the substrate comes to close the peptide and form a cycle.

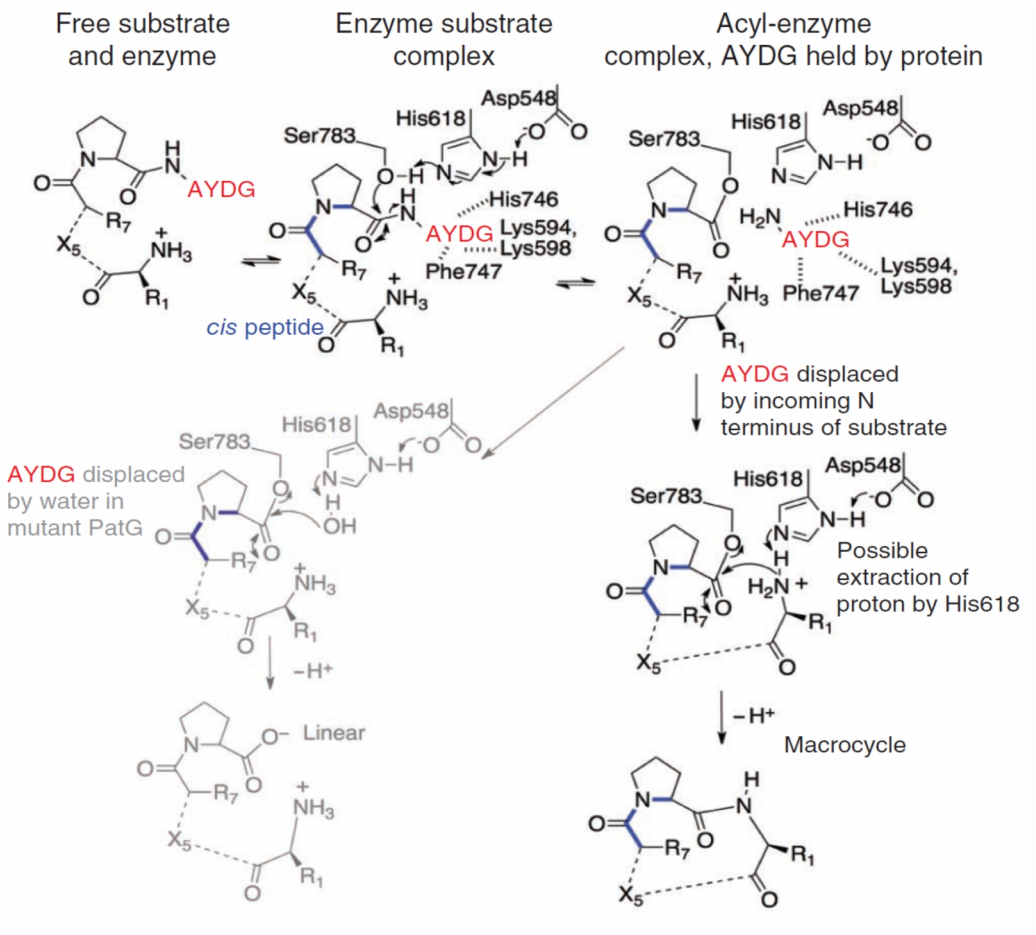


Fig.23 – Proposed mechanism for macrocyclization [82].

4. MODELLING OF THE SYSTEM

To date, the structure of the PatG macrocyclase domain (359 residues) have solved and its complex with a substrate mimic (12 residues, 9 resolved), highlighting the key interactions which enable macrocyclisation over standard proteolysis [77]. Hence, we started the modelling from this experimentally obtained complex of the enzyme and a substrate mimic (PDB code: 4AKT), and free enzyme PatGmac (PDB code: 4AKS). The great difficulty was to model the native and correct substrate inside the enzyme. Since the

X-Ray structure contains inside the enzyme a mimic propeptide not yet subjected to the microcylasation and epimerisation, and we need a correct substrate. The natural PatD propeptide has the following sequence: Ile-Thr-Ala-Cys-Ile-Thr-Phe-Cys Ala-Tyr-Asp-Gly. In the experiment, the mimic peptide has the sequence Val-Gly-Ala-Gly-Ile-Gly-Phe-Pro Ala-Tyr-Asp-Gly. Four residues Ala-Tyr-Asp-Gly in both substrates are recognition signal residues. Presumably, proline mimics the five-membered ring of the heterocycle in the natural product. In Fig. 24 the X-ray structure (a), and the system with correct substrate (b) after modelling are presented.

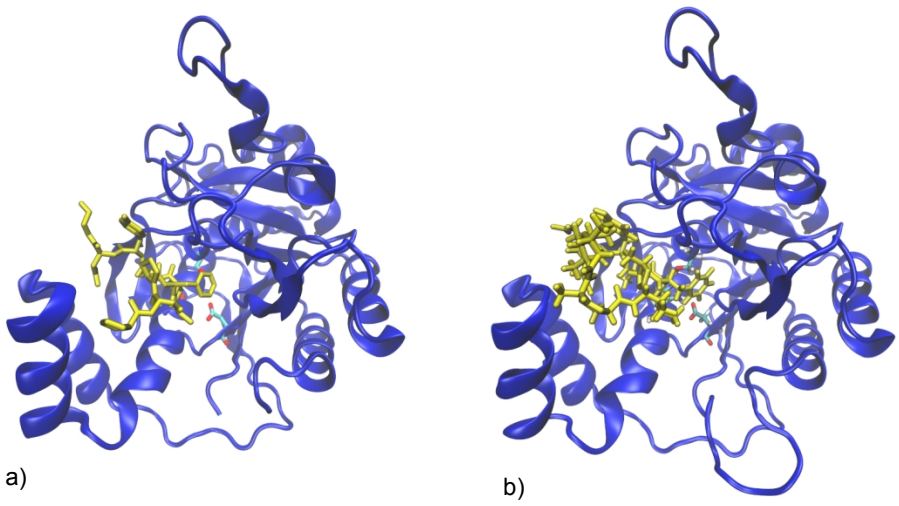


Fig. 24 – The representation of: a) the X-ray structure of the substrate-enzyme complex; b) the corrected structure of the substrate-enzyme complex.

First, we started from the correction of the active site. In the X-ray structure, the His618 of the active-site triad in the substrate-enzyme complex was substituted by Ala. However, in the free enzyme the correct residues were used. The structure of the free enzyme was superimposed with the structure of the complex, and the coordinates of the atoms of the residue of Ala618 were replaced by the coordinates of the atoms of the correct residue His618 in the complex.

After that, the loop (residues 651 – 657), missing in X-ray structure, was modelled in MODELLER [35].

Then, we started with correcting the substrate of the complex. The idea was to conserve the position of the most of the residues in the substrate and model missing parts. In the experiment, the peptide not yet converted to microcylisation was used. We were interesting in the system of the complex substrate-enzyme with already converted to oxidation and epimerization changes. Besides, in the experiment was used initially not correct substrate – prolines were used to mimic heterocycles, and the epimerisation centres were not correct. We started from taking off the coordinates of the substrate. After

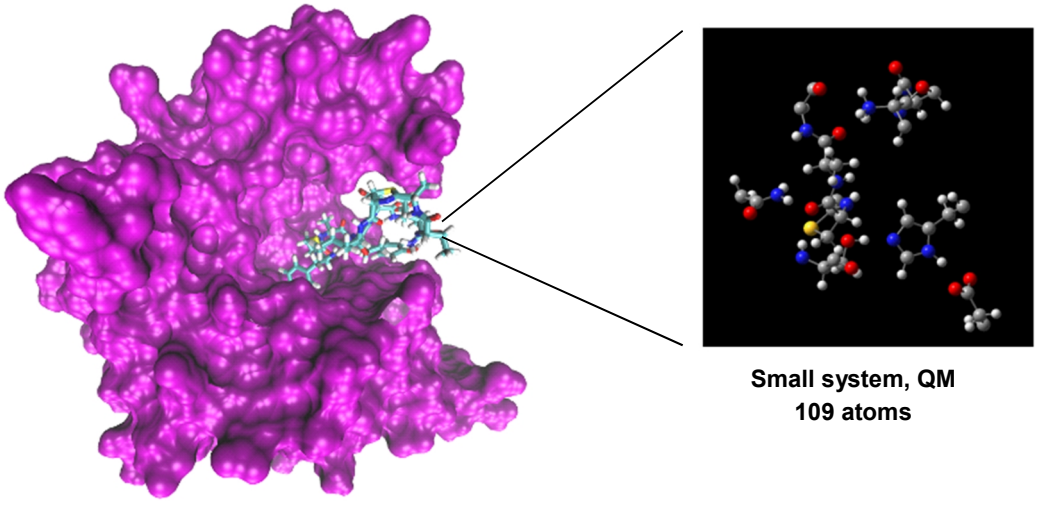
correcting the heterocycles, and modelling in the software Gausview [85] the missing parts, we transfer back the substrate to the corrected enzyme. After that, the whole system were minimized in Amber in 3 steps, in order to relax the system, removed the tensions and clashes of the new substrate, and adding the loop and correction of the active site.

5. QM/MM OPTIMIZATION

When the system was optimized in Amber, we proceed with the optimization in ONIOM QM/MM.

The QM/MM method is a hybrid quantum mechanical/molecular mechanics method. This scheme permits to divide the system into different regions that can be studied with different level of theory such as *ab initio*, DFT, semi-empirical and molecular mechanics, to produce reliable geometries and energies at reduced computational time.

In ONIOM QM/MM the energy of the full model system is calculated at the MM level. Then the energy of the small model (the region of interest) is subtracted at the MM level and added at the QM level (Fig. 25).



E(ONIOM) = E(MM, big) – E(MM, small) + E(QM, small)

Fig. 25 – ONIOM model of the system.

The whole system includes 5200 atoms. Into the high layer we put the active site residues, the place where the substrate is going to be cleaved, and the terminal part of the peptide that comes to close the macrocycle. In total 109 atoms were considered (Fig. 25).

After the full optimization of the system, detailed examination of the system was done. The modelled and optimized structure shows appropriate position of the active site residues and the substrate (Fig. 26). Thus, Ser783 is in a geometrically beneficial position to the substrate to start a nucleophilic attack of the carbonyl. The distance between the atoms is 3.72 Å. His618 is close and in good orientation to proceed with the next step of the catalytic mechanism (the distance is 1.82 Å). And, finally, Asp548 is close enough to His618 with the distance of 1.94 Å.

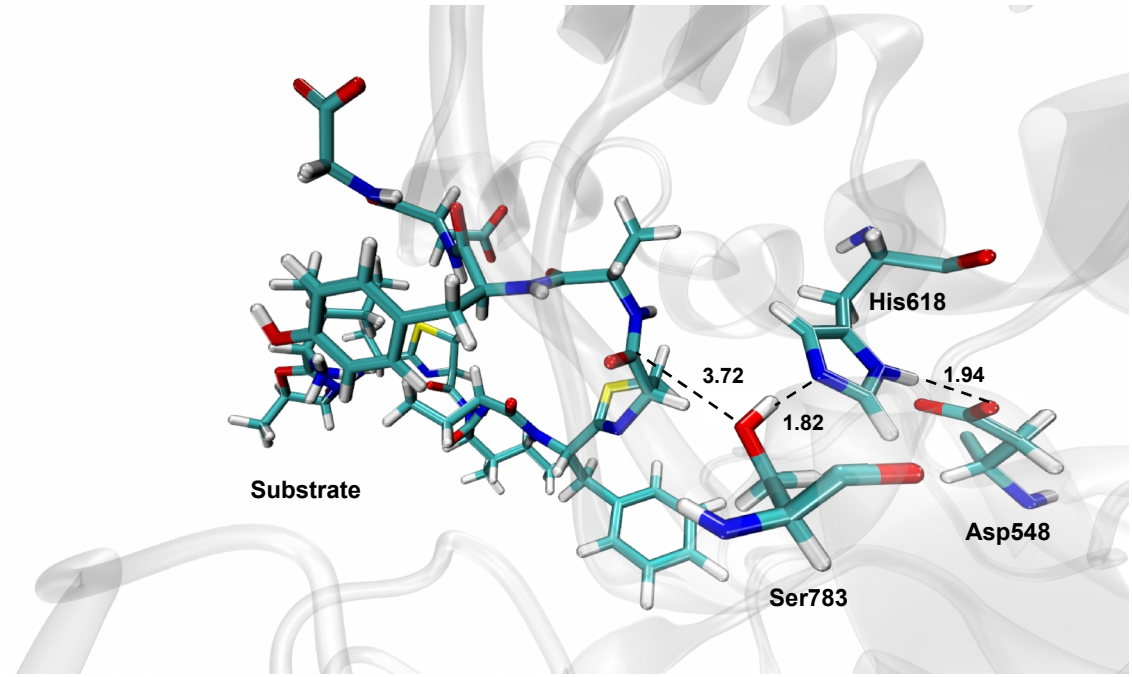


Fig. 26 – The representation of substrate in the active site of the enzyme after the optimisation.

After this, the system is prepared for the study of the catalytic mechanism using QM/MM method.

6. PROSPECTIVE FOR THE FUTURE

During this project the modelling of the system of enzyme and correct substrate was done with the help of computational technics in order to use this system in the study of catalytic mechanism with the atomistic level of theory. The study of the catalytic mechanism can be done in three steps: nucleophilic attack of Ser of the carbonyl atom, proton transfer from Ser to His, and proton transfer from His to Asp. These three steps are followed by cleavage of the four stop signalling residues, and approaching of the N-termini to the cleaved part to close and form a macrocycle of patellamide.

III. IDENTIFYING PROTEIN STRUCTURE WITH INFRARED SPECTROSCOPY

1. INTRODUCTION AND MOTIVATION

Two-dimensional infrared spectroscopy (2DIR) [86, 87], applied to the amide-I vibrations (C=O stretch) of proteins, has been shown to offer a powerful tool for the study of protein conformational change. Since the very first experimental demonstration of the technique, numerous proteins have been studied experimentally with the technique [88, 89] and ways of distinguishing typical secondary structures as α-helices [90], β-sheets [89], and 3^{10}-helices [91] have been developed. Only within the last few years theoretical methods capable of detailed simulation of two-dimensional spectra of full proteins from molecular dynamics trajectories have become available [92, 93]. Still these new methods have not been widely tested on available experimental data. The goal of the present project is to apply these new simulation methods to proteins for which the experimental data are readily available (as myoglobin, lysozyme, ribonuclease A, and concanavalin A [89]) in order to benchmark the calculation methods.

In this project, molecular dynamics simulations of the proteins of interest have been performed using the GROMACS MD-package [94]. From these trajectories the needed vibrational frequencies and transition dipoles have been extracted [95 – 97] using three different schemes. This information has then been processed with the newest version of the NISE program [92, 98] to calculate the 2DIR spectra of the different proteins. The spectra have been compared with experimental spectra, obtained from the Tokmakoff group at the University of Chicago [99], and analysed on the base of visual observation, overlap and RMSD between experimental and simulated spectra.

2. 2D INFRARED SPECTROSCOPY OF PROTEINS

Two-dimensional infrared spectroscopy is a new approach in the study of protein conformation and conformational dynamics that occur on picosecond to millisecond time scales. As a vibrational spectroscopy, it examines the vibrations of chemical bonds and how many vibrations of a molecule and its environment interact with one another. However, macromolecules as proteins have so many degrees of freedom that FTIR spectra are often considered ambiguous therefore they display little variation for different

proteins due to the broad and featureless line shape that arises from different structural motifs. One can simplify the visualization of complex spectra consisting of many overlapped peaks by spreading a vibrational spectrum over second independent frequency dimension to reveal two-dimensional line shapes and cross-peaks. Cross-peaks are hallmark of multidimensional spectroscopy. They are a measure of the coupling between molecular vibrations and thus contain information on the molecular structure.

Infrared absorption is a result of the interaction between an electromagnetic field and the oscillating molecular dipole moment. In one-dimensional experiment there is one independent time or frequency variable. For example, in a frequency-domain experiment, one measures the change of light intensity through a sample as a function of the frequency of a monochromatic field. In the case of two-dimensional spectroscopy, one measures the change to the absorption spectrum induced by excitation at an independent frequency. This also can be performed in the frequency-domain as a double resonance experiment with independent excitation and detection beams. Fourier transform methods are commonly applied in one or both dimensions of the 2D spectrum. Following the initial short pulse excitation (Fig. 27), subsequent pulses can interact with the sample polarization in order to generate a nonlinear polarization. Since multiple pulses can excite/de-excite the sample multiple times, the final nonlinear polarization radiates light carrying frequencies with new information about the molecular potential not present in linear absorption spectrum.

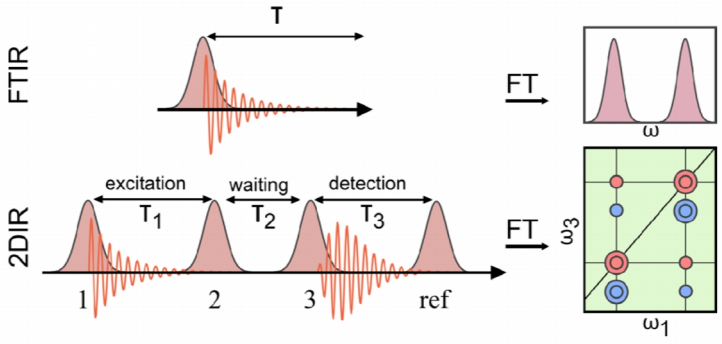


Fig. 27 – Pulse sequences for FTIR and 2D IR spectroscopy [100].

Structural information is extracted from the position and intensities of the peaks, whereas dynamics are extracted from the peak shapes, more specifically, the ellipticity of the peaks as well as the change in ellipticity with waiting time. Exciting a molecule during the first time delay and probing it during the third time delay, the observation of the signal of the molecule would be at a point in the spectrum connected with the frequency it had during the first delay along the ω_1 axis, and the frequency it had during the third time delay along the ω_3 axis. If the frequency didn't change during waiting time then it would be

observed a diagonal elongation, referred to as inhomogeneous broadening. If on the other hand the dynamics is very fast compared to the waiting time all molecules will have lost their memory of their initial frequency after the waiting time and the peak will be round, or elongated antidiagonal, referred to as homogeneous broadening.

In the very first experimental demonstration of the technic 2D IR was applied to the amide-I vibrations that are known for secondary structure sensitivity. We review 2D IR spectroscopy of the amide I protein backbone vibration.

Modelling of Amide I Spectroscopy

Amide group vibrations of the protein backbone are of great attention in protein IR spectroscopy because they are native to all proteins and the absorption is strong. The most studied amide I modes (1600 – 1700 cm^{-1}) that include primarily C=O stretching are delocalized over the entire protein backbone, and are sensitive to the global structure of the protein. Amide I vibrations are largely isolated from the other protein vibrations and can be described as linear combinations of local amide I vibrations at each peptide unit. With the exception of proline the side chain vibrations do not interact strongly with the amide I vibration.

The sensitivity of amide I vibrations to the secondary structure of the protein resulting from strong coupling between amide I oscillator of different residues and the coupling depends on the structure of the protein. The peaks for α-helices and unstructured regions are observed at 1640–1660 cm^{-1} and 1640–1650 cm^{-1}, respectively. While for β-sheets two peaks are observed, a strong absorption band near 1630–1640 cm^{-1} and a weaker band at high frequencies (>1680 cm^{-1}) [100, 101].

Model for amide I spectroscopy

On a computational level a theoretical model for amide I spectroscopy must account for *local* and *global effects*.

Local effects come from the influence of an oscillator's (local modes are often referred to as oscillators or sites) local electrostatic environment on its vibrational frequency (also known as site energy). It is the influence of local interactions on bond strength within the amide unit. For example, hydrogen bonds between water molecules and oxygen of the amide unit stabilizes the C=O bond, lowering its vibrational frequency (and causing a red-shift of the corresponding amide I absorption peak). Therefore, this site can be associated with the local hydrogen bonding configuration in which the peptide resides. The correlation of local structure and vibrational frequency can be expressed in terms of electrostatic variables (potential, field, gradient, etc.), which is the basis for most of the site energy maps. Although the details of these methods vary, all seek to predict IR

spectroscopic features based on the local electrostatic environment of each oscillator in a given structure.

Global effects come from the delocalization of amide excitations due to its site-to-site coupling. For description of the vibrational frequencies of individual sites local electrostatic effects would be sufficient. But interactions between neighbouring residues induce delocalization of vibrational modes across multiple residues, thus given rise to a more complicated picture. For this reason, beyond electrostatics, the central feature of most theoretical descriptions of amide I vibrations is the concept of the vibrational exciton – a delocalized excited state produced by the interaction of many individual sites. These vibrational excitons observed in amide I spectra can be compared to a network of oscillators – or sites – coupled together into a single system. Vibrational coupling between individual amide units in a protein causes excitation of any given amide vibration to induce vibrations in adjacent units. This delocalization of vibrational motion over a system depends on the number of sites, the site-to-site coupling strength, and the variation of site energies.

3. PROCEDURE

A mixed quantum-classical approach is used in the methods to model FTIR and 2DIR spectra. The principal assumption is an adiabatic separation of the amide I vibrational motions and the dynamics of the protein. It is stated that essentially all large amplitude protein motions, such as conformational fluctuations, can be treated classically via static protein structures or MD simulations, while the higher frequency vibrations, such as amide I which are examined by the infrared radiation, must be treated quantum-mechanically. The interaction between the amide I vibrations and protein structure is handled with the mapping procedures.

The amide I Hamiltonian that describes the interactions between N sites includes doubly excited states. The exciton states are described as eigenstates of the following exciton Hamiltonian (equation 1):

$$\begin{aligned}\hat{H} &= \sum_{n=1}^{N} \varepsilon_n \left|n\right\rangle\left\langle n\right| + \sum_{m,n=1}^{N} J_{mn} \left|m\right\rangle\left\langle n\right| \\ &+ \sum_{m,n=1}^{N} (\varepsilon_m + \varepsilon_n - \Delta\delta_{mn}) \left|mn\right\rangle\left\langle mn\right| \\ &+ \sum_{m,n=1}^{N} \sum_{\substack{j,k=1 \\ (m,n)\neq(j,k)}}^{N} J_{mn,jk} \left|mn\right\rangle\left\langle jk\right|\end{aligned} \tag{1}$$

Here, ε_n and J_{mn} represent the site energies and coupling constants between singly-excited states. The second set of terms represent doubly-excited states, in which a

single oscillator is excited twice (*m*=*n*), or two different oscillators are each excited (*m*≠*n*). Δ is the anharmonicity value (~16 cm^{-1} for amide I) due to the anharmonicity of harmonic oscillator. This value is particularly important for 2D spectroscopy. This value is the difference in the absorption frequency for the fundamental transition (0 → 1) compared to with its overtone (1 → 2). For a perfectly harmonic system (with Δ=0), the 2D IR signal vanishes due to interference between the fundamental and overtone transitions. The coupling between the one- and two-quantum states is neglected, so that the states are separated into zero, one-, and two-quantum subspaces. The resulting block-diagonal amide I Hamiltonian is illustrated graphically in Fig. 28, in which site energies occur along the diagonal and site-to-site couplings appear off-diagonal within the one- and two-exciton blocks. The eigenstates of the system are obtained by numerically diagonalizing the resulting matrix, providing absorption frequency values, dipole moments, and oscillator strengths. Usually, a harmonic approximation is assumed to obtain the two-exciton coupling constants $J_{mn,jk}$ from the one-exciton energies ε_n and coupling constants J_{mn}.

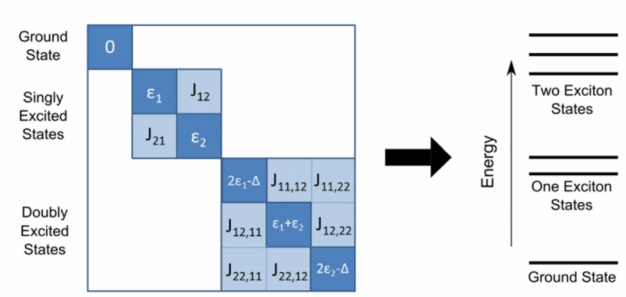


Fig. 28 – Schematic representation of the excitonic Hamiltonian matrix and energy levels for a 2-oscillator system [100].

To calculate those values, it is necessary to know the site energies and coupling constants for each state. These parameters can be obtained from molecular structures (for example, from MD simulations). To extract them from the MD simulation, a variety of parametrized maps is used, without the need for electronic structure calculations. The maps correlate the local structure (site energies are dependent on the local structure) and electrostatic environment of the amide bond with the frequencies of the associated vibrational modes in order to greatly reduce computational expense of calculating site frequencies.

The maps are constructed from electronic structure calculations on *N*-methyl acetamide (NMA), a model system for the peptide bond monomer. The maps assume a linear correlation between the frequency and electrostatic parameters. For the study in the present project were selected three following maps: Jansen, Skinner and Tokmakoff maps.

Jansen map was parameterized through DFT calculations on NMA – a model of peptide bond – in point charge environment [95].

Skinner map [102] is an empirical map, constructed on the base of NMA in different solvents.

Tokmakoff map is an empirical map, constructed on the base of Alanine and Glycine dipeptides [103].

The second set of parameters – the site-to-site coupling constants J_{mn} was first modelled using the transition dipole coupling (TDC) model [104, 105]. TDC model treats through-space coupling by treating each oscillator as a simple dipole vector which interacts with adjacent dipoles through a coupling factor. An improvement model of the TDC is the transition charge coupling (TCC) [95, 106] model in which a transition charge is assigned to each atom of the amide bond, and interaction energy is calculated between each atom. In the present study the TCC model for the couplings between amide I vibrations of different sites was used.

After amide-I Hamiltonian is constructed, it's diagonalized and the eigenstates and eigenvectors are used to calculate transition frequencies and dipole moments. These values are then used for simulating spectra.

The procedure for simulating linear absorption and 2D IR spectra can be summarised in the following scheme (Fig. 29). The Hamiltonian is constructed by combining a MD trajectory with a mapping procedure to obtain the amide I frequencies and couplings. After that, scheme of solving the Shrödinger equation is applied to obtain the spectra.

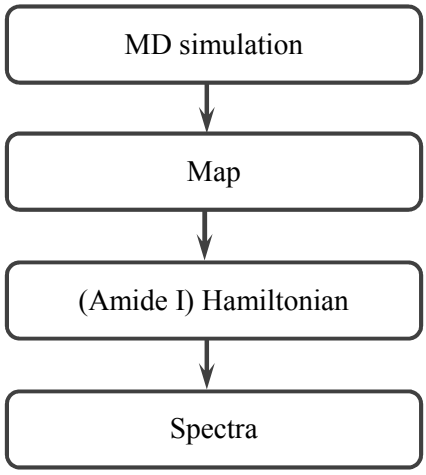


Fig. 29 – A diagram for the procedure of simulating linear absorption and 2D IR spectra

4. COMPUTATIONAL DETAILS

Protein selection

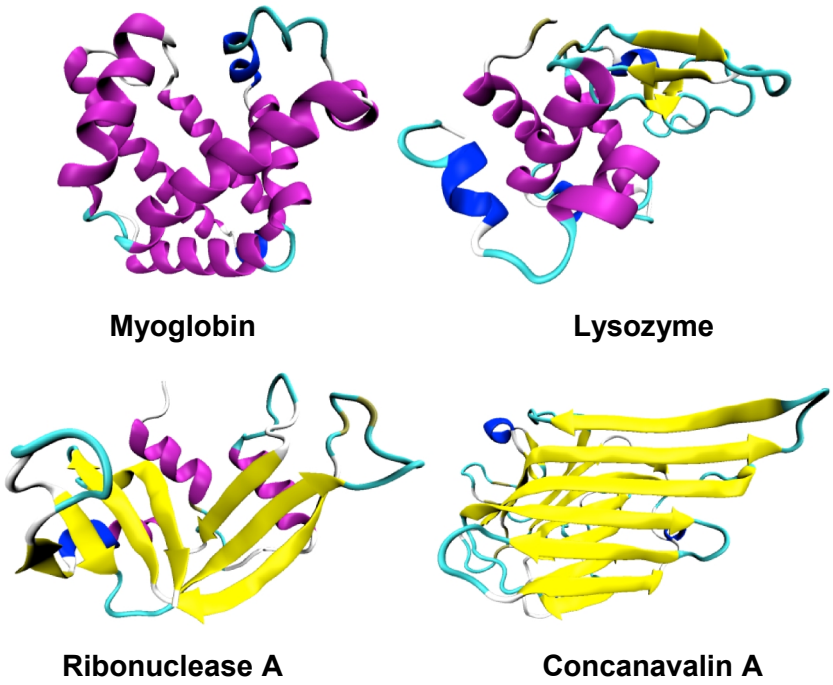


Fig. 30 – The structures (in ribbon representation) of the four proteins under the study.

The set of proteins used in the study (Fig. 30) was selected to include a variety of structures ranging from mainly α-helical protein to mainly β-sheet protein. There were chosen four proteins for which experimental data had already been obtained by Tokmakoff group [99]. Myoglobin (Mb, 153 residues, PDB ID 5MBN; 0% β-sheet, 74% α-helix structure) has no β-sheets. Lysozyme (Lys, 129 residues, PDB ID 1AKI; 6% β-sheet, 31% α-helix structure) has a small three-stranded β-sheet region. Ribonuclease A (RNse A, 124 residues, PDB ID 1FS3; 32% β-sheet, 18% α-helix structure) has two domains that vary from two- to four-stranded regions. And Concanavalin A (Con A, 237 residues, PDB ID 1NLS; 46% β-sheet, 0% α-helix structure is the most extended system, with two relatively flat six-stranded AP β-sheets.

MD simulations

MD simulations were performed with GROMACS-4.6.5 [94]. Except for myoglobin, for all the proteins the implemented improved side-chain torsion potentials for the Amber ff99SB protein force field – Amber 99SB-ILDN [107] was used. Due to the presence of parameters for heme group of myoglobin, MD simulation of this protein was performed using GROMOS 54a7 [108], where the parameters are implemented. The protein

crystallographic structures were downloaded from the Protein Data Base. Missing protons were filled in, and the protein structures were solvated in explicit water solvent SPC/E water [109], and chlorine or sodium ions were added to keep the system neutral. Finally, the systems were relaxed (to be sure that no steric clashes or inappropriate geometry take place) by energy minimization with 50000 steps.

Equilibration was conducted in two phases: 100-ps NVT equilibration (constant Number of particles, Volume, and Temperature) that stabilized the temperature of the system, and 100-ps NPT equilibration (Number of particles, Pressure, and Temperature are all constant) that stabilizes the pressure (and thus also the density) of the system.

Upon completion of the two equilibration phases, the system was well-equilibrated at the desired temperature and pressure. Afterwards a 1-ns MD simulation with 2 ps time step was performed. The structures and trajectories were stored every 20 fs, giving a total of 50 000 stored snapshots.

Constructing of the amide-I Hamiltonian

For every snapshot of MD trajectory, the amide-I Hamiltonian was constructed, using AmideImaps v. 1.0.0, to calculate transition dipole moments and site frequencies. For frequency calculation, three different schemes for electrostatic maps were used: Jansen map, Skinner map and Tokmakoff map.

Long range couplings, i.e. couplings between non-nearest neighbours, were treated with TCC. Nearest neighbour Couplings, i.e. between nearest neighbours, were treated with GLDP scheme, by the Torii and Tasumi map [111], based on the Hartree-Fock ab-initio calculations of the glycin dipeptide (GLDP) for different Ramachandran angles.

Spectral simulation (NISE v. 2A)

The linear absorption and 2D IR spectra were calculated using the NISE method [98, 110], which is based on numerically solving the time-dependent Schrödinger equation for the amide I Hamiltonian.

In the spectral simulation, the side-chain carbonyl vibrations were neglected (aspartic acid, glutamic acid, asparagine, and glutamine units). The frequency of this vibration is in the same range as that of the backbone carbonyls, but no reliable model for the solvent dependence of its frequency exists.

5. RESULTS, ANALYSIS AND DISCUSSION

5.1 Obtained linear spectra

Experimental FTIR spectra are shown in Fig. 31 [99]. Qualitatively, absorption spectra associated with primarily α-helical proteins myoglobin and lysozyme are characterized by a single band centred near 1650 cm^{-1} with an approximate diagonal width of 50 cm^{-1}, while proteins composed of primarily β-sheet show two peaks centred near 1620-1640 and 1680 cm^{-1}, respectively, resulting from vibrations whose main transition dipoles lie perpendicular and parallel to the β-strands respectively.

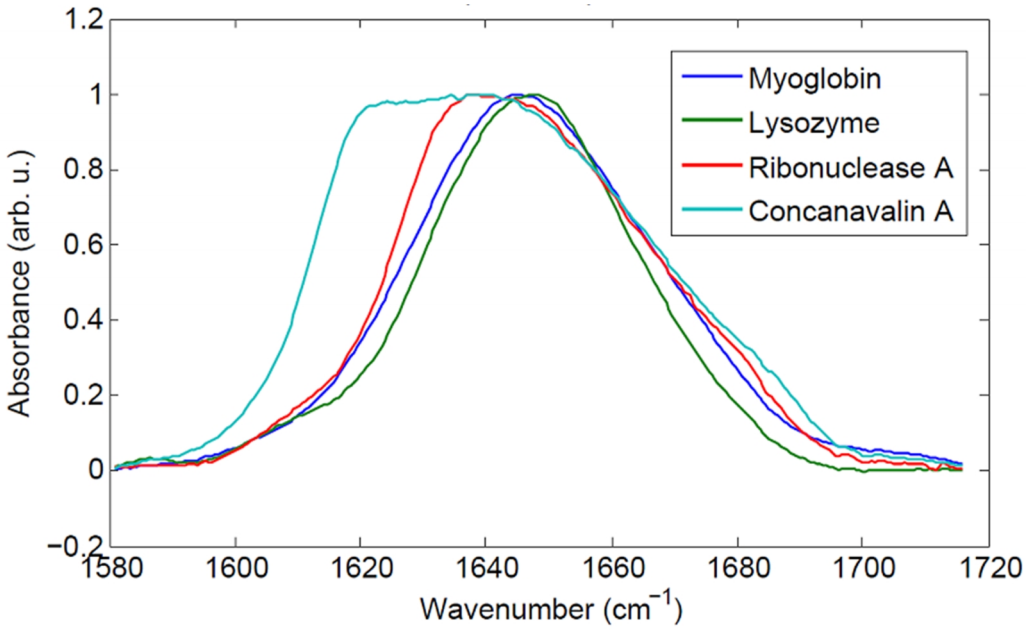


Fig. 31 – Experimental FTIR spectra of the proteins.

The FTIR spectra show that, going from exclusively α-helix system myoglobin to exclusively β-sheet system Con A, the peak amide-I transition frequency red-shifts from 1650 cm^{-1} and the absorption lines change their shape from symmetric to more decline-shaped, and the peaks are becoming broader. This trend is expected with increasing β-sheet content. Nevertheless, it is difficult to correlate the spectral profiles observed in FTIR spectra with different structural elements present in this series of four proteins. In the FTIR spectra, the two-peak structure indicating the presence of a β-sheet is hardly evident. Only a weak shoulder is observed for Con A and RNse A, while no such shoulder is present for myoglobin and lysozyme.

Theoretical linear absorption spectra using different maps are shown in the Fig. 32. For clear visual comparison, the peak positions of the predicted spectra were moved to the peak positions of experimental spectra. The shifts for each protein and each map are presented in the Table 7. Excluding shift peak positions of myoglobin, for which another force filed was used for MD simulations, there are some dependencies. Thus,

Jansen scheme regularly overestimates frequencies for about 11 cm^{-1}. Skinner and Tokmakoff schemes underestimate frequencies, and the underestimation values is less regular, for about 15 and 13cm^{-1} for Skinner and Tokmakoff maps, respectively. These blue- and red-shifting frequencies should be done while applying corresponding map.

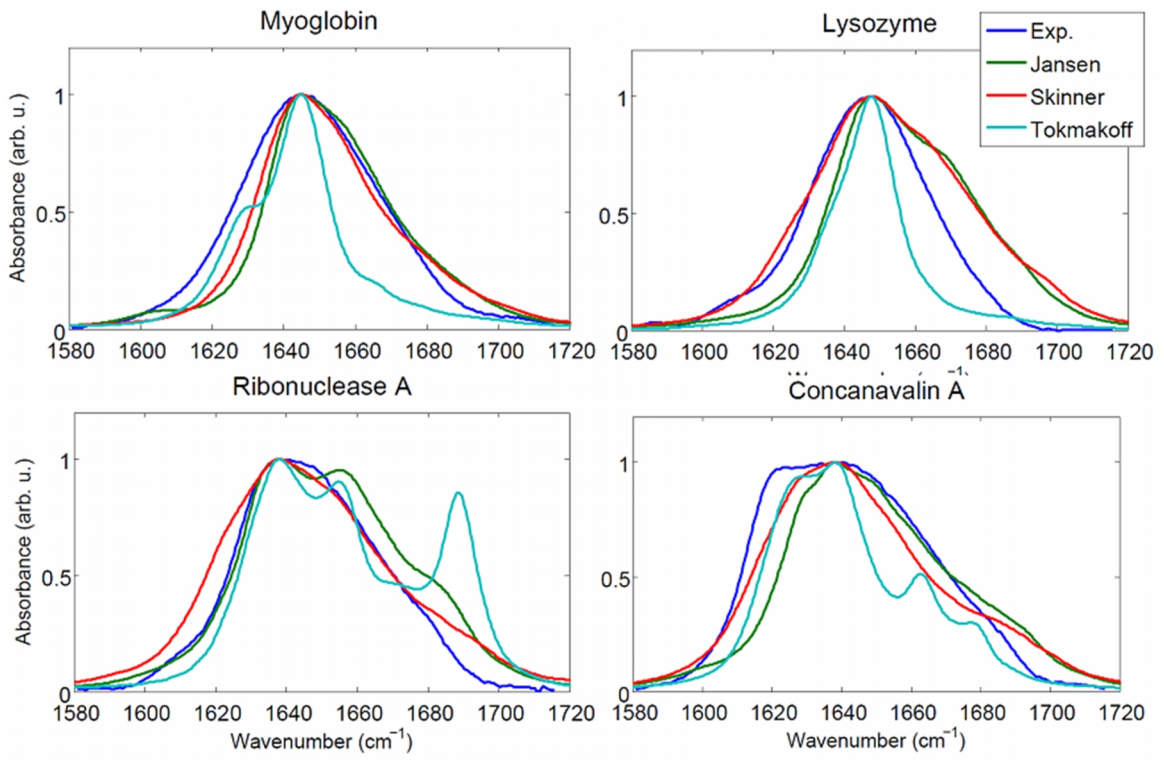


Fig. 32 – Predicted FTIR spectra for the proteins simulated using Jansen, Skinner and Tokmakoff maps. Simulated spectra are shifted to the peak position of experimental spectra.

	Mb	Lys	RNse A	Con A
Jansen	-3.8	-7.4	-12.1	-12.2
Skinner	21.9	19.9	11.4	12.7
Tokmakoff	23.5	10.9	18.4	10.2

Table 7 – Shift of the peak positions.

Looking onto spectra one can notice that generally Tokmakoff map gives narrower spectra in comparison with the experimental spectra. As well for the β-sheet structures RNse A and Con A there are multiple peaks. These can be a result of narrower distribution of site frequencies. The analyses of the average site frequency distribution (Table 8) showed that for Tokmakoff map the deviation from the average frequency is the smallest.

		Mb	Lys	RNse A	Con A
Jansen	av. frequency (cm^{-1})	1665.6	1670.1	1671.5	1671.0
	av. deviation (cm^{-1})	17.7	19.1	21.4	22.9
Skinner	av. frequency (cm^{-1})	1638.8	1639.4	1643.5	1642.9
	av. deviation (cm^{-1})	18.8	23.0	26.1	26.0
Tokmakoff	av. frequency (cm^{-1})	1660.6	1667.5	1662.8	1661.8
	av. deviation (cm^{-1})	17.7	19.6	20.1	19.9

Table 8 – Frequency site distribution of the protein.

To measure the difference between values predicted by models and the values actually observed in the experiment the root-mean-square deviation (RMSD) was used, Equation 2.

$$RMSD = \sqrt{\sum (\mathrm{y}_{exp} - \mathrm{y}_{theor})^2} \tag{2}$$

Analysing the values of RMSD presented below in the Table 9, it is clear that the best values are observed for Skinner map, and the worst for Tokmakoff map.

	Mb	Lys	RNse A	Con A
Jansen	2.023	2.911	1.696	2.832
Skinner	1.479	2.490	1.550	1.629
Tokmakoff	3.234	3.123	3.346	2.939

Table 9 – RMSD analysis between theoretical and experimental spectra.

To measure the similarity of the spectra quantitatively, the value of overlapping *S* (Equation 3) between each simulated spectra and experimental one was used, as well as overlapping between experimental spectra of proteins.

$$S = \frac{y_{exp} \cdot y_{theor}}{\sqrt{\sum (\mathrm{y}_{exp}^2 - \mathrm{y}_{theor}^2)}} \tag{3}$$

The results of overlapping are presented in the diagram below (Fig. 33). For estimation of these values two limits were used, presented by blue and red horizontal lines. These limits can be used to exclude inconsistent results. Thus, the lowest value 0.9449 is a value of overlapping between the two most different structures in this subset of proteins revealed in the experiment; therefore under this limit theoretical models are not able to differentiate between the two proteins. The highest value 0.9886 is a value of overlapping between the two most similar structures; therefore it is likely a lot to have overlapping between theoretical and experimental spectra close or higher to this limit. As can be seen from this diagram the best values are obtained for the Skinner map, where

the overlapping in most cases is close to the upper limit and for Con A it's even higher. The Jansen map, except for Lysozyme, as well showed high overlapping. Whereas, the worst values are obtained for the Tokmakoff map.

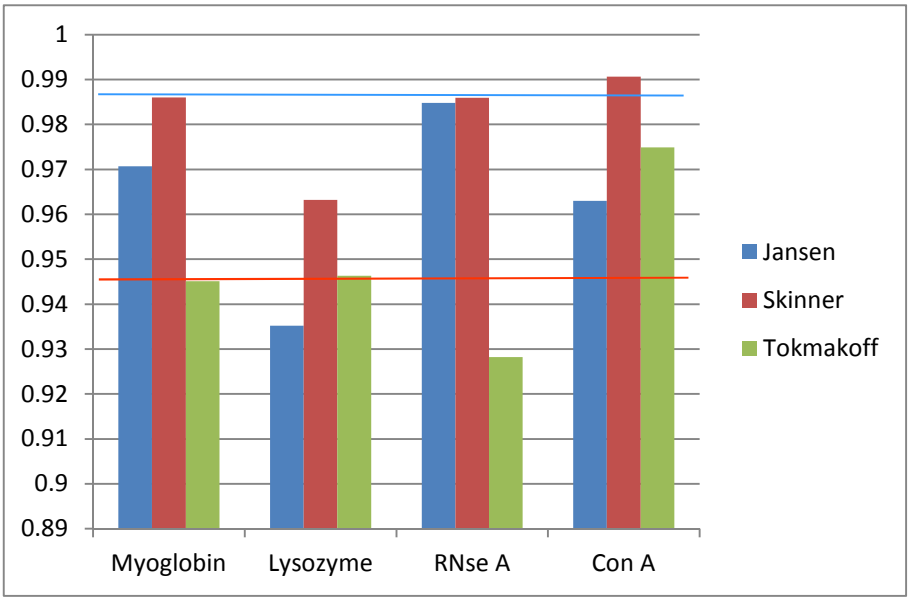


Fig. 33 – Diagram of overlapping between simulated FTIR spectra and experimental ones for each protein. Upper limit is the highest overlapping between experimental spectra of proteins. Lower limit is the lowest overlapping between experimental spectra of proteins.

5.2. Obtained 2DIR spectra

Experimentally obtained 2D IR spectra [99] are given in the Fig. 34. The 2D IR spectra show considerably more features about the structure in comparison with linear absorption and allow the β-sheet features not apparent in the FTIR spectra to be discriminated.

In proteins, interference effects between diagonal and cross-peaks, combined with contributions from disordered regions and other secondary structures, lead to a "Z"-shaped contour profile in the 2D IR amide I correlation spectrum that is characteristic indicator of the presence of AP β-sheet secondary structure.

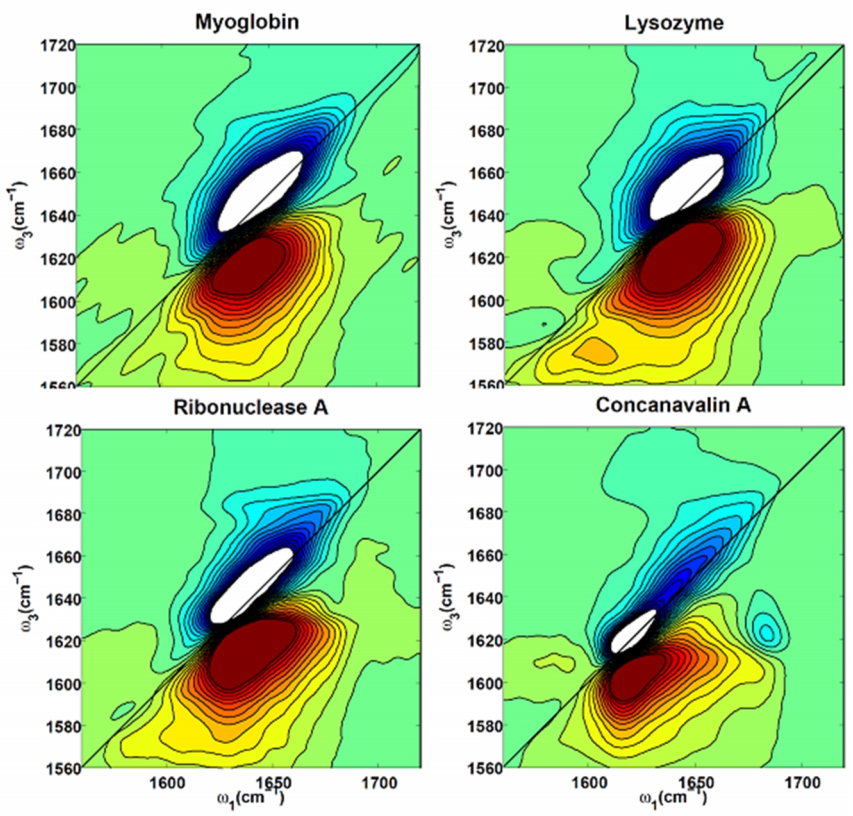


Fig. 34 – Experimental 2DIR spectra for the proteins.

In the Fig. 35 are presented theoretical 2D IR spectra together with experimental spectra (on the top of the figure). For Con A all maps reproduce cross-peak, related to the β-sheet secondary structure. It can be observed as broadening of the peak.

All the peaks and shoulders, appeared in the linear absorption spectra, are highlited in the 2D spectra. Thus, there are clear peaks in the Tokmakoff map for β-sheet structures of RNse A and Con A. There are present diagonal peaks in the spectra of myoglobin, corresponding to the shoulders in the linear absorption spectra.

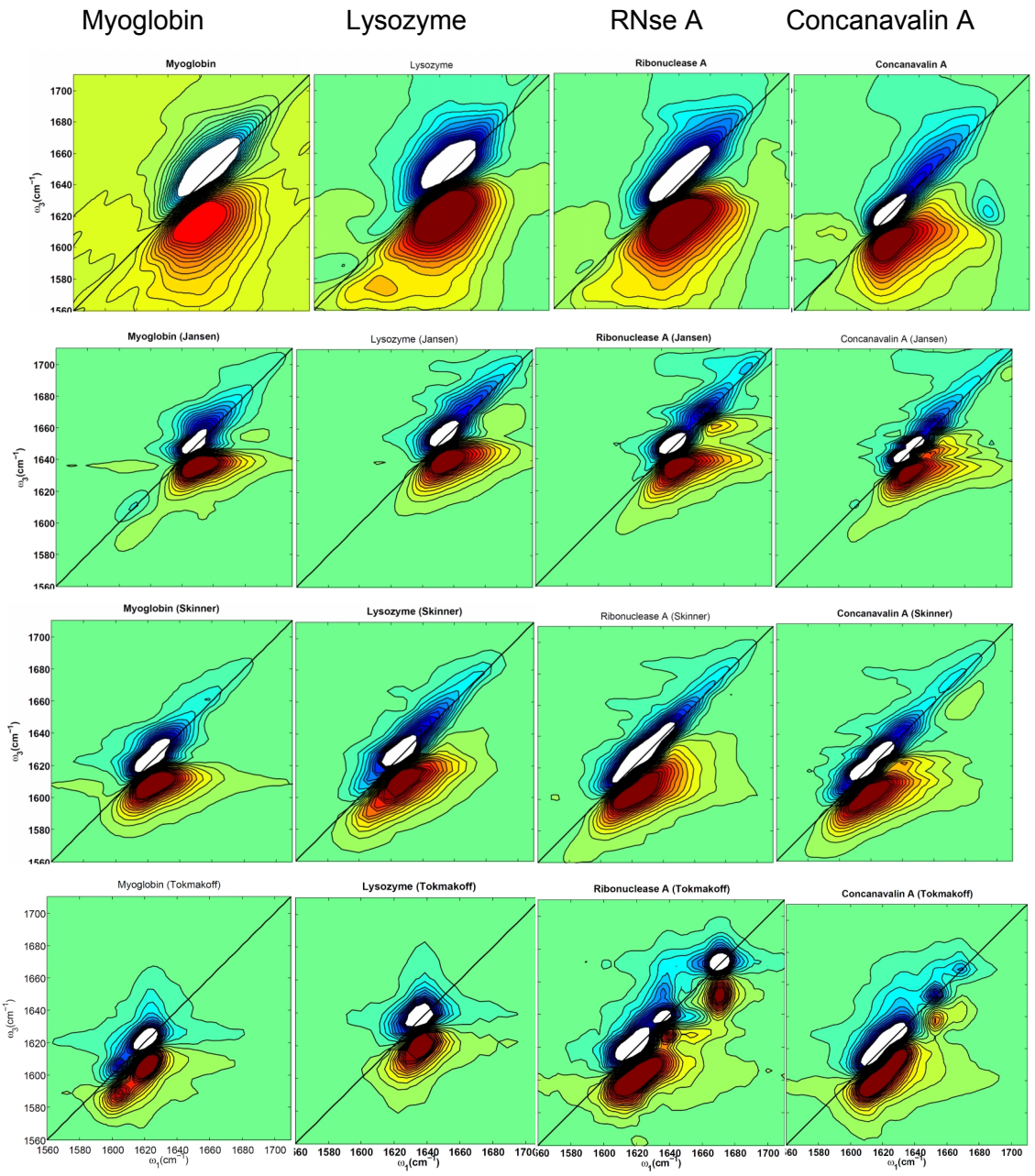


Fig. 35 – Experimental and simulated spectra for the proteins.

In the Table 10 are presented the values of RMSD. The values are bigger than the ones for linear spectra, but the tendency is the same – Skinner map gives slightly better results.

	Mb	Lys	RNse A	Con A
Jansen	25.803	28.482	28.441	23.800
Skinner	26.754	27.140	22.146	10.515
Tokmakoff	30.757	27.580	27.729	12.214

Table 10 – RMSD analysis between theoretical and experimental 2D IR spectra.

The same as for linear spectra, overlapping analysis was done. The overlapping between experimental and theoretical spectra was done for each protein and each map without moving peak positions. The results of overlapping are presented in the diagram below (Fig. 36). From the diagram it is clearly seen that the Skinner map gives slightly higher values.

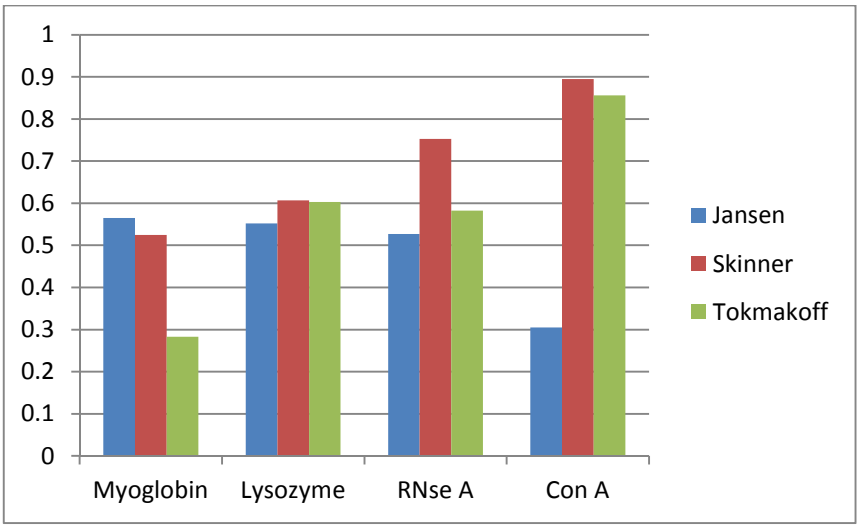


Fig. 36 – Diagram of overlapping between simulated and experimental 2D IR spectra.

In general, Tokmakoff map is worst in reproducing the shapes of the experimental spectra, while Jansen and Skinner maps are considerably close to each other. But the values of overlapping doesn't show the worst results, probably due to the contribution of small additional peaks to the overlapping. Better results showed Skinner map, reproducing better the shap of the peaks, and having the smalles RMSD values and highest values of overlapping.

6. CONCLUSIONS AND OUTLOOK

The results show that 2D IR spectroscopy is more sensitive to structural differences between proteins than traditional infrared spectroscopy. Infrared spectroscopy is thus an appealing probe because it is intrinsic to all proteins and is sensitive to conformation structure. Moreover, as a method with intrinsic time resolution in the picosecond range, infrared methods can be applied to study the fast kinetics of protein conformational changes accompanying protein folding.

Although there are models for simulating 2D spectra of proteins, there are still difficulties that can be faced. One of them is absorption of side-chain vibrations in the amide I region. Despite the fact that several side-chain groups absorb in the amide I spectral region, these moieties are usually neglected for spectral calculations, while the spectra of real peptides will contain contributions from side-chain vibrations in the amide I

region. An accurately parameterized site energy map for these groups would be of interest for many applications, particularly since their vibrations are expected to be largely localized, making them potentially useful as probes of local structure. Benchmarking of such parameterized site energy maps is very useful since it gives an opportunity to estimate the given method using particular set of parameters applying for particular systems. Together with estimating errors in calculations it helps identifying opportunities for improvement the method. This is useful for developing new mappings.

In this project the calculation of 2D spectra was done, using one set of parameters. In principle, one may vary and adjust the parameters to his own need. For example, by testing on other systems, changing force field for MD simulation, comparing coupling models.

REFERENCES

[1] R.M. Carey, AT1 Receptors, Angiotensin Receptor Blockade, and Clinical Hypertensive Disease; In *Renin Angiotensin System and Cardiovascular Disease Contemporary Cardiology*; W.C. DeMello, E.D. Frohlich, Eds.; *Humana Press*, **2009**:59.

[2] R.M. Carey, Pathophysiology of Primary Hypertension; In *Handbook of Physiology: Microcirculation*; R.F. Tuma, W.N. Duran, K. Ley, Eds.; *Elsevier: Amsterdam*, **2008**, 2:794.

[3] J. Drews, Drug Discovery: A Historical Perspective; *Science*, **2000**, 287:1960.

[4] X. Jeunemaitre, A.P. Gimenez-Roqueplo, J. Célérier, P. Corvol, Angiotensinogen variants and human hypertension; *Curr. Hypertens. Rep.*, **1999**, 1:31.

[5] A. Cappelli, G.L. Mohr, A. Gallelli, M. Rizzo, M. Anzini, S. Vomero, L. Mennuni, F. Ferrari, F. Makovec, M.C. Menziani, P.G. De Benedetti, G. Giorgi, Design, Synthesis, Structural Studies, Biological Evaluation, and Computational Simulations of Novel Potent AT1 Angiotensin II Receptor Antagonists Based on the 4-Phenylquinoline Structure; *J. Med. Chem.*, **2004**, 47:2574.

[6] B. Scheiper, H. Matter, H. Steinhagen, U. Stilz, Z. Bocskei, V. Fleury, G. McCort, Discovery and optimization of a new class of potent and non-chiral indole-3-carboxamide-based renin inhibitors; *Bioorg. Med. Chem. Lett.*, **2010**, 20:6268.

[7] C. Jensen, P. Herold and H.R. Brunner, Aliskiren: the first renin inhibitor for clinical treatment; *Nat. Rev. Drug Discov.*, **2008**, 7:399.

[8] G.K. Aulakh, R.K. Sodhi and M. Singh, An update on non-peptide angiotensin receptor antagonists and related RAAS modulators; *Life Sciencies*, **2007**, 81:615.

[9] F. Messerli, M.A. Weber, H.R. Brunner, Angiotensin II receptor inhibition: A new therapeutic principle; *Arch. Intern. Med.*, **1996**, 156:1957.

[10] T.L. Goodfriend, M.E. Elliott, K.J. Catt, Angiotensin receptors and their antagonists; *N. Engl. J. Med.*, **1996**, 334:1649.

[11] M. Caffrey, Membrane Protein Crystallization; *J. Struct. Biol.*, **2003**, 142:108.

[12] E. Krieger, S.B. Nabuurs, and G. Vriend, Homology Modeling; In *Structural Bioinformatics*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2003**, 44.

[13] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, Crystal structure of rhodopsin: A G protein-coupled receptor; *Science*, **2000**, 289:739.

[14] V. Cherezov, D.M. Rosenbaum, M.A. Hanson, S.G. Rasmussen, F.S. Thian, T.S. Kobilka, H.J. Choi, P. Kuhn, W.I. Weis, B.K. Kobilka, R.C. Stevens, High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor; *Science*, **2007**, 318(5854): 1258.

[15] D. Wacker, G. Fenalti, M.A. Brown, V. Katritch, R. Abagyan, V. Cherezov, R.C. Stevens, Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography; *J. Am. Chem. Soc.*, **2010**, 132(33):11443.

[16] B. Wu, E.Y. Chien, C.D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F.C. Bi, D.J. Hamel, P. Kuhn, T.M. Handel, V. Cherezov, R.C. Stevens, Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists; *Science*, **2010**, 330(6007):1066.

[17] M. Burnier, Angiotensin II type 1 receptor blockers; *Circulation*, **2001**, 103:904.

[18] R.K. Turker, I.H. Page, F.M. Bumpus, Antagonists of angiotensin II; In: *Handbook of Experimental Pharmacology*, I.H. Pagi, F.M. Bumpus, Eds.; *Angiotensin*, Springer Verlag, New York, **1974**, 37:162.

[19] D.H.P. Streeten, G.H. Anderson, J.M. Freiberg, T.G. Dalakos, Use of an angiotensin II antagonist (saralasin) in the recognition of "angiotensinogenic" hypertension; *N. Engl. J. Med.*, **1975**, 292:657.

[20] Y. Furakawa, S. Kishimoto, K. Nishikawa, Hypotensive imidazole derivatives and hypotensive imidazole-5-acetic acid derivatives; *U.S. Patents* 4,340,598 and 4,355,040. Osaka, Japan, **1982**.

[21] E.K. Jackson, Renin and angiotensin; In: L. Brunton, J. Lazo, K. Parker, et al., Eds., *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, 11th Ed. New York: McGraw-Hill, **2006**:789.

[22] P.B. Timmermans, P.C. Wong, A.T. Chiu, W.F. Herblin, P. Benfield, D.J. Carini, R.J. Lee, R.R. Wexler, J.A. Saye, R.D. Smith, Angiotensin II receptors and angiotensin II receptor antagonists; *Pharmacol. Rev.*, **1993**, 45:205.

[23] P. Buhlmayer, P. Furet, L. Criscione, M. de Gasparo, S. Whitebread, T. Schmidlin, R. Lattmann, J. Wood, Valsartan, a potent, orally active angiotensin II antagonist developed from the structurally new amino acid series; *Bioorganic & Medicinal Chemistry Letters*, **1994**, 4:29.

[24] B. Christophe, R. Libon, C. Cazaubon, D. Nisato, A. Manning, P. Chatelain, Effects of irbesartan (SR 47436/BMS-186295) on angiotensin IIinduced pressor responses in pithed rat: potential mechanisms of action; *European Journal of Pharmacology*, **1995**, 281:161.

[25] Y. Shibouta, Y. Inada, M. Ojima, T. Wada, M. Noda, T. Sanada, K. Kubo, K. Nishikawa, Pharmacological profile of a highly potent and long-acting angiotensin II receptor antagonist, 2-ethoxy-1-[[2'-(1H-tetrazol-5-yl)biphenyl-4-yl]methyl]-1H-benzimidazole-7-carboxylic acid (CV-11974), and its prodrug, (+/-)-1-

(cyclohexyloxycarbonyloxy)-ethyl-2-ethoxy-1-[[2'-(1H-tetrazol-5-yl)biphenyl-4-yl]methyl]-1H-benzimidazole-7-carboxylate (TCV-116); *Journal of Pharmacology and Experimental Therapeutics*, **1993**, 266:114.

[26] K. Nishikawa, Y. Inada, Y. Shibouta, T. Wada, M. Ojima, K. Kubo, T. Naka, Pharmacological profile of a novel nonpeptide angiotensin II subtype 1 receptor antagonist, TCV-116; *Blood Pressure Suppl.*, **1994**, 5:7.

[27] M. Noda, Y. Shibouta, Y. Inada, M. Ojima, T. Wada, T. Sanada, K. Kubo, Y. Kohara, T. Naka, K. Nishikawa, Inhibition of rabbit aortic angiotensin II (AII) receptor by CV-11974, a new nonpeptide AII antagonist; *Biochemical Pharmacology*, **1993**, 46:311.

[28] T.L. Blundell, S. Bedarkar, E. Rinderknecht, R.E. Humble, Insulin-like growth factor: a model for tertiary structure accounting for immunoreactivity and receptor binding; *Proc Natl Acad Sci.*, **1978**, 75:180.

[29] I.T. Weber, Evaluation of homology modeling of HIV protease; *Proteins*, **1990**, 7(2):172.

[30] C.B. Anfinsen, Principles that govern the folding of protein chains, *Science*, **1973**, 181:223.

[31] M.S. Johnson, N. Srinivasan, R. Sowdhamini, T.L. Blundell, Knowledge-based protein Modelling; *Crit Rev Biochem Mol Biol.*, **1994**, 29(1):1.

[32] J.M. Chandonia, S.E. Brenner, Implications of structural genomics target selection strategies: Pfam5000, whole genome, and random approaches; *Proteins*, **2005**, 58:166.

[33] D. Vitkup, E. Melamud, J. Moult, C. Sander, Completeness in structural genomics; *Nat Struct Biol.*, **2001**, 8:559.

[34] Modeller: https://salilab.org/modeller/about_modeller.html

[35] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints; *J. Mol. Biol.*, **1993**, 234:779.

[36] B. John, A. Sali, Comparative protein structure Modelling by iterative alignment, model building and model assessment; *Nucleic Acids Res*, **2003**, 31(14):3982.

[37] J. Wang, P. Cieplak, P. Kollman, How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *Journal of Computational Chemistry*, **2000**, 21(12):1049.

[38] M. Takeda-Shitaka, D. Takaya, C. Chiba, H. Tanaka, H. Umeyama, Protein structure prediction in structure based drug design; *Curr Med Chem*, **2004**, 11:551.

[39] K. Joo, J. Lee, J Lee, Methods for accurate homology modeling by global optimization; *Methods Mol Biol*., **2012**, 857:175.

[40] E. Shakhnovich, Protein folding thermodynamics and dynamics: Where physics, chemistry, and biology meet; *Chem. Rev.*, **2006**, 106:1559.

[41] C.L. Brooks, M. Karplus, B.M. Pettitt, Proteins: A theoretical perspective of dynamics, structure, and thermodynamics; *Wiley*, New York, **1988**:xiii.

[42] S. Tanaka, H.A. Scheraga, Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins; *Macromolecules*, **1976**, 9:945.

[43] M.Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures; *Protein Sci.*, **2006**, 15(11):2507.

[44] F. Melo, R. Sanchez, A. Sali, Statistical potentials for fold assessment; *Protein Science*, **2002**, 11:430.

[45] R. Henderson, J.M. Baldwin, T.A. Ceska, F. Zemlin, E. Beckmann, K.H. Downing, Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy; *J Mol Biol*, **1990**, 213:899.

[46] M.P. Joseph, B. Maigret, J.C. Bonnafous, J. Marie, H.A. Scheraga, A computer modeling postulated mechanism for angiotensin II receptor activation; *J. Protein Chem.*, **1995**, 14:381.

[47] Y. Yamano, K. Ohyama, M. Kikyo, T. Sano, Y. Nakagomi, Y. Inoue, N. Nakamura, I. Morishima, D.F. Guo, T. Hamakubo, T. Inagami, Mutagenesis and the molecular modeling of the rat angiotensin II receptor (AT1); *J. Biol. Chem.*, **1995**, 270:14024.

[48] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank; *Nucleic Acids Research*, **2000**, 28(1):235.

[49] D.C. Teller, T. Okada, C.A. Behnke, K. Palczewski, R.E. Stenkamp, Advances in Determination of a High-Resolution Three-Dimensional Structure of Rhodopsin, a Model of G-Protein-Coupled Receptors (GPCRs); *Biochemistry*, **2001**, 40:7761.

[50] J. Li, P.C. Edwards, M. Burghammer, C. Villa, G.F. Schertler, Structure of bovine rhodopsin in a trigonal crystal form; *J Mol Biol.*, **2004**, 343(5):1409.

[51] T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, E.M. Landau, Y. Shichida, Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography; *Proc. Natl. Acad. Sci. USA*, **2002**, 99:5982.

[52] T. Okada, M. Sugihara, A.N. Bondar, M. Elstner, P. Entel, V. Buss, The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure; *J. Mol. Biol.*, **2004**, 342:571.

[53] H. Wu, D. Wacker, V. Katritch, M. Mileni, G.W. Han, E. Vardy, W. Liu, A.A. Thompson, X.-P. Huang, F.I. Carroll, S.W. Mascarella, R.B. Westkaemper, P.D. Mosier, B.L. Roth, V. Cherezov, R.C. Stevens, Structure of the human kappa opioid receptor in complex with JDTic; *Nature*, **2012**, 485(7398):327.

[54] SwissProt Database: http://www.uniprot.org/

[55] ClustalW2: http://www.ebi.ac.uk/Tools/msa/clustalw2/

[56] J.A. Ballesteros, H. Weinstein, Integrated methods for the construction of three dimensional models and computational probing of structure function relations in G protein-coupled receptors; *Methods Neurosci.*, **1995**, 25:366.

[57] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of Polypeptide Chain Configurations; *J Mol Biol*, **1963**, 7:95.

[58] A.L. Morris, M.W. MacArthur, E.G. Hutchinson, J.M. Thornton, Stereochemical quality of protein structure coordinates; *Proteins*, **1992**, 12:345.

[59] Amber: http://ambermd.org/

[60] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility; *J Comput Chem.*, **2009**, 30(16):2785

[61] N.M.F.S.A. Cerqueira, J. Ribeiro, P.A. Fernandes, M.J. Ramos, vsLab—An implementation for virtual high-throughput screening using AutoDock and VMD; *Int. J. Quantum Chem.*, **2011**, 111:1208

[62] F.L.P. Fierens, P.M.L. Vanderheyden, Z. Gáborik, T. Le Minh, J.P. De Backer, L. Hunyady, A. Ijzerman, G. Vauquelin. Lys 199 mutation of the human angiotensin type 1 receptor differentially affects the binding of surmountable and insurmountable non-peptide antagonists; *J. Renin-Angiotensin-Aldosterone System*, **2000**, 1:283.

[63] P. Zoumpoulakis, T. Mavromoustakos, Seeking the active site of the AT1 receptor for computational docking studies; *Drug Design Reviews-Online*, **2005**, 2:537.

[64] L. Hunyady, H. Ji, G. Jagadeesh, M. Zhang, Z. Gaborik, B. Mihalik, K.J. Catt, Dependence of AT(1) Angiotensin receptor function on adjacent asparagine residues in the 7th transmembrane helix; *Molecular pharmacology*, **1998**, 54(2):427.

[65] H. Ji, W. Zheng, Y. Zhang, K.J. Catt, K. Sandberg, Genetic transfer of a nonpeptide antagonist binding site to a previously unresponsive angiotensin receptor; *PNAS*, **1995**, 92(20):9240.

[66] T. Takezako, C. Gogonea, Y. Saad, K. Noda, S.S. Karnik, “Network leaning” as a mechanism of insurmountable antagonism of the angiotensin II type 1 receptor by non-peptide antagonists; *J. Biol. Chem.*, **2004**, 279:15248.

[67] L. Yan, B.J. Holleran, P. Lavigne, E. Escher, G. Guillemette, R. Leduc, Analysis of transmembrane domains 1 and 4 of the human angiotensin II AT1 receptor by cysteine-scanning mutagenesis; *J. Biol. Chem.*, **2010**, 285:2284.

[68] K. Noda, Y. Saad, A. Kinoshita, T.P. Boyle, R.M. Graham, A. Husain, S.S. Karnik, Tetrazole and carboxylate groups of angiotensin receptor antagonists bind to the same subsite by different mechanisms; *J Biol Chem*, **1995**, 270:2284.

[69] H.T. Schambye, S.A. Hjorth, J. Weinstock, T.W. Schwartz, Interaction between the nonpeptide angiotensin antagonist SKF-108,566 and histidine 256 (HisVI:16) of the angiotensin type 1 receptor; *Mol Pharmacol*, **1995**, 47(3):425.

[70] H.M. Han, S.I. Shimuta, C.A. Kanashiro, L. Oliveira, S.W. Han, A.C. Paiva, Residues Val254, His256, and Phe259 of the angiotensin II AT1 receptor are not involved in ligand binding but participate in signal transduction; *Mol Endocrinol*, **1998**, 12(6):810.

[71] S.S. Martin, B.J. Holleran, E. Escher, G. Guillemette, R. Leduc, Activation of the angiotensin II type 1 receptor leads to movement of the sixth transmembrane domain: Analysis by the substituted cysteine accessibility method; *Mol. Pharmacol.*, **2007**, 72:182.

[72] D. Fillion, J. Cabana, G. Guillemette, R. Leduc, P. Lavigne, E. Escher, Structure of the human angiotensin II type 1 (AT1) receptor bound to angiotensin II from multiple chemoselective photoprobe contacts reveals a unique peptide binding mode; *J. Biol. Chem.*, **2013**, 288:8187.

[73] ChemSpider: http://www.chemspider.com/

[74] Drug Bank: http://www.drugbank.ca/

[75] PubMed database: http://www.ncbi.nlm.nih.gov/pubmed

[76] ZINC database: http://zinc.docking.org/

[77] J. Koehnke, A. Bent, W.E. Houssen, D. Zollman, F. Morawitz, S. Shirran, J. Vendome, A.F. Nneoyiegbe, L. Trembleau, C.H. Botting, M.C.M. Smith, M. Jaspars, J.H. Naismith, The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain; *Nat Struct Mol Biol.*, **2012**, 19(8):767.

[78] B.F. Milne, P.F. Long, A. Starcevic, D. Hranueli, M. Jaspars, Spontaneity in the patellamide biosynthetic pathway; *Org Biomol Chem.*, **2006**, 4(4):631.

[79] D.F. Sesin, S.J. Gaskell, C.M. Ireland, The Chemistry of Lissoclinum Patella; *Bull. Soc. Chim. Belges*, **1986**, 95:853.

[80] E.W. Schmidt, J.T. Nelson, D.A. Rasko, S. Sudek, J.A. Eisen, M.G. Haygood, J. Ravel, Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella; *Proc. Natl. Acad. Sci. USA*, **2005**, 102:7315.

[81] P.F. Long, W.C. Dunlap, C.N. Battershill, M. Jaspars, Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production; *ChemBioChem*, **2005**, 6:1760.

[82] E.W. Schmidt, The hidden diversity of ribosomal peptide natural products; *BMC Biol.*, **2010**, 8:83.

[83] W.E. Houssen, M. Jaspars, Azole-based cyclic peptides from the sea squirt Lissoclinum patella: old scaffolds, new avenues; *ChemBioChem*, **2010**, 11:1803.

[84] J.A. McIntosh, C.R. Robertson, V. Agarwal, S.K. Nair, G.W. Bulaj, E.W. Schmidt, Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst; *J. Am. Chem. Soc.*, **2010**, 132:15499.

[85] Gaussian: http://www.gaussian.com/g_prod/gv5.htm

[86] P. Hamm, M.H. Lim, R.M. Hochstrasser, Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy; *J. Phys. Chem. B*, **1998**, 102:6123.

[87] P. Hamm, M.T. Zanni, Concepts and Methods of 2D Infrared Spectroscopy; *Cambridge University Press*, Cambridge, **2011**.

[88] Z. Ganim, K.C. Jones, A. Tokmakoff, Insulin dimer dissociation and unfolding revealed by amide I two-dimensional infrared spectroscopy; *Chem. Phys.*, **2010**, 12:3579.

[89] N. Demirdöven, C.M. Cheatum, H.S. Chung, M. Khalil, J. Knoester, A. Tokmakoff, Two-Dimensional Infrared Spectroscopy of Antiparallel β-Sheet Secondary Structure; *J. Am. Chem. Soc.*, **2004**, 126:7981.

[90] S. Woutersen, P. Hamm, Time-resolved two-dimensional vibrational spectroscopy of a short α-helix in water; *J. Chem. Phys.*, **2001**, 115:7737.

[91] H. Maekawa, C. Toniolo, A. Moretto, Q.B. Broxterman, N.-H. Ge, Different spectral signatures of octapeptide 3^{10}- and α-helices revealed by two-dimensional infrared spectroscopy; *J. Phys. Chem. B*, **2006**, 110:5835.

[92] C. Liang, T.L.C. Jansen, An efficient N3-Scaling propagation scheme for simulating two-dimensional infrared and visible spectra; *J. Chem. Theory Comput.*, **2012**, 8:706.

[93] C. Liang, M. Louhivuori, S.J. Marrink, T.L.C. Jansen, J. Knoester, Vibrational spectra of a mechanosensitive channel; *J. Phys. Chem. Lett.*, **2013**, 4:448.

[94] E. Lindahl, B. Hess, D.v.d. Spoel, GROMACS 3.0: A package for molecular simulation and trajectory analysis; *J. Mol. Model.*, **2001**, 7:306.

[95] T.L.C. Jansen, J. Knoester, A transferable electrostatic map for solvation effects on amide I vibrations and its application to linear and two-dimensional spectroscopy; J. *Chem. Phys.*, **2006**, 124:044502.

[96] T.L.C. Jansen, A.G. Dijkstra, T.M. Watson, J.D. Hirst, J. Knoester, Modeling the amide I bands of small peptides; *J. Chem. Phys.*, **2006**, 125:044312.

[97] S. Roy, J. Lessing, G. Meisl, Z. Ganim, A. Tokmakoff, J. Knoester, T.L.C. Jansen, Solvent and conformation dependence of amide I vibrations in peptides and proteins containing proline; *J. Chem. Phys.*, **2011**, 135:234507.

[98] T.L.C. Jansen, J. Knoester, Nonadiabatic Effects in the two-dimensional infrared spectra of peptides: Application to alanine dipeptide; *J. Phys. Chem. B*, **2006**, 110:22910.

[99] C.R. Baiz, C.S. Peng, M.E. Reppert, K.C. Jones, A. Tokmakoff, Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution; *Analyst*, **2012**, 137:1793.

[100] C.R. Baiz, M. Reppert, A. Tokmakoff, Introduction to Protein 2D IR Spectroscopy. Ultrafast Infrared Vibrational Spectroscopy; Michael D. Fayer, Ed.; *CRC Press*, **2013**, 12.

[101] Z. Ganim, H.S. Chung, A.W. Smith, L.P. Deflores, K.C. Jones, A. Tokmakoff, Amide I two-dimensional infrared spectroscopy of proteins; *Accounts of chemical research*, **2008**, 41(3):432.

[102] L. Wang, C.T. Middleton, M.T. Zanni, J.L. Skinner, Development and validation of transferable amide I vibrational frequency maps for peptides; *J. Phys. Chem. B*, **2011**, 115(13):3713.

[103] M. Reppert, A. Tokmakoff, Electrostatic frequency shifts in amide I vibrational spectra: direct parameterization against experiment; *J. Chem. Phys.*, **2013**, 138:134116.

[104] S. Krimm, J. Bandekar, Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins; *Adv Protein Chem*, **1986**, 38:181.

[105] H. Torii, M. Tasumi, Model calculations on the amide-I infrared bands of globular proteins; *J. Chem. Phys*, **1992**, 96(5):3379.

[106] P. Hamm, S. Woutersen, Coupling of the amide I modes of the glycine dipeptide; *Bull. Chem. Soc. Jpn.*, **2002**, 75:985.

[107] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field; *Proteins*, **2010**, 78:1950.

[108] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W.F. van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7; *Eur Biophys J.*, **2011**, 40(7):843.

[109] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials; *J. Phys. Chem.*, **1987**, 91:6269.

[110] T.L.C. Jansen, J. Knoester, Simulation of two-dimensional infrared spectra by numerical integration of the Schrödinger equation. In *International Conference of Computational Methods in Sciences and Engineering*, Recent Progress in Computational Sciences and Engineering, Vol. 7. T. Simos, G. Maroulis, Eds.; *Brill Academic Publishers*, Leiden, The Netherlands, China, Crete, Greece, **2006**

[111] H. Torii, M. Tasumi, Ab initio molecular orbital study of the amide I vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix; *J. Raman Spec.*, **1998**, 29(1):81.

APPENDIX

Theoretical spectra of proteins for each tested map